PROFESSOR ADRIANO NUNES-NESI (Orcid ID : 0000-0002-9581-9355)
DR THOMAS GUIRAUD (Orcid ID : 0000-0003-3254-5307)
DR ALISDAIR R FERNIE (Orcid ID : 0000-0001-9000-335X)
DR MATHILDE CAUSSE (Orcid ID : 0000-0002-0407-4985)
PROFESSOR FERNANDO CARRARI (Orcid ID : 0000-0003-0145-156X)

Article type : Original Article

Validated MAGIC and GWAS populations mapping reveal the link between vitamin E contents and natural variation in chorismate metabolism in tomato

Author information

Estanislao Burgos[1], Maria Belen De Luca[1], Isidore Diouf[2], Luis A. de Haro[1], Elise Albert[2,3], Christopher Sauvage[2,4], Zhao Jian Tao[2], Luisa Bermudez [5,6], Ramon Asís[7], Adriano Nunes Nesi[8], Michel Matringe[9], Claire Brehelin[10], Thomas Guiraud[9], Carine Ferrand[9], Isabelle Atienza[9], Joana Jorly[9], Jean Philippe Mauxion[9], Pierre Baldet[9], Alisdair R. Fernie[12], Leandro Quadrana[5,13], Christophe Rothan[9], Mathilde Causse[2], Fernando Carrari[1]

Affiliations

[1]Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE-UBA-CONICET), Universidad de Buenos Aires, Ciudad Universitaria, C1428EHA Buenos Aires, Argentina.

[2]INRAE, GAFL, 84143, Monfavet, France

[3]Present address: Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, United States.

[4]Present address: Syngenta Seeds SAS, 12 Chemin de l'Hobit, 31790 Saint Sauveur, France.

[5]Instituto de Biotecnología, Instituto Nacional de Tecnología Agropecuaria and Consejo Nacional de Investigaciones Científicas y Técnicas, PO Box 25, B1712WAA, Castelar, Argentina.

[6]Facultad de Agronomía. Cátedra de Genética. Universidad de Buenos Aires. Argentina.
[7]CIBICI, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, CC 5000, Córdoba, Argentina.
[8]Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, Minas Gerais 36570-900, Brazil.
[9]Laboratoire de Physiologie Cellulaire Végétale, Unité Mixte de Recherche 5168 CNRS-CEA-INRA, Université Joseph Fourier, CEA Grenoble, 38054 Grenoble Cedex 9,PCV, CEA-Grenoble, 38054, Grenoble, France.
[10]UMR 1332 Biologie du Fruit et Pathologie, INRA, Univ. Bordeaux, 33140, Villenave d'Ornon, France.
[11]CNRS, Laboratory of Membrane Biogenesis, UMR 5200, 33140 Villenave d´Ornon Cedex, France.
[12]Max-Planck-Institut für Molekulare Pflanzenphysiologie, Wissenschaftspark Golm, Am Mühlenberg 1, 14476 Potsdam, Germany.
[13]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), F-75005 Paris, France.
Email addresses:
Estanislao Burgos <burgosestanislao@gmail.com>
Maria Belen De Luca < bdeluca@fbmc.fcen.uba.ar>
Isidore Diouf < dioufisi@hotmail.com>
Luis A. de Haro <luideharo@gmail.com>
Elise Albert <alberte2@msu.edu>
Christopher Sauvage <christopher.sauvage.1@gmail.com>
Zhao Jiantao < jnzhao_cornell@163.com>
Luisa Bermudez <bermudez.luisa@inta.gob.ar>
Ramon Asís <ramonasis@gmail.com>
Adriano Nunes Nesi <nunesnesi@ufv.br>
Michel Matringe <mmatringe@cea.fr>
Claire Brehelin <claire.brehelin@inrae.fr>
Thomas Guiraud <thomas@guiraud.co>
Carine Ferrand <carine.ferrand@inrae.fr>

Isabelle Atienza <isabelle.atienza@inrae.fr>
Joana Jorly <joana.jorly@inrae.fr>
Jean Philipppe Mauxion <jean-philippe.mauxion@inrae.fr>
Pierre Baldet <pierre.baldet@inrae.fr>
Alisdair R. Fernie <Fernie@mpimp-golm.mpg.de>
Leandro Quadrana <quadrana@biologie.ens.fr>
Christophe Rothan <christophe.rothan@inrae.fr>
Mathilde Causse <mathilde.causse@inrae.fr>

Address correspondence: Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE-UBA-CONICET), Universidad de Buenos Aires, Ciudad Universitaria, C1428EHA Buenos Aires, Argentina.
To whom correspondence should be sent. e-mail: **carrarifernando@gmail.com** and **mathilde.causse@inrae.fr**

Telephone: (+54-11) 4576-3368/3386/3428 int. (269).

Running title: Functional validation of genetic analysis in vitamin E

Key words: Tocopherol, Tocotrienol, Vitamin E, Tomato Fruit, *Solanum lycopersicum*.

Summary

Tocochromanols constitute the different forms of vitamin E (VTE), essential for the human diet and display a high membrane protectant activity. By combining interval mapping and genome-wide association studies (GWAS) we unveiled the genetic determinants of tocochromanol accumulation in tomato fruits. To enhance the nutritional value of this highly consumed vegetable, we dissected the natural intraspecific variability of tocochromanols in tomato fruits and genetically engineered their biosynthetic pathway. These analyses allowed the identification of a total of 25 QTL interspersed across the genome pinpointing the chorismate-tyrosine pathway as a regulatory hub controlling the supply of the aromatic head group for tocochromanol biosynthesis. To validate the link between the chorismate-tyrosine pathway and VTE, we engineered tomato plants to bypass the pathway at the arogenate branchpoint. Transgenic tomatoes showed moderate increments in tocopherols (up to ~ 20%) and a massive accumulation of tocotrienols (up to ~ 3,400%). Gene expression analyses of these plants reveal a trade-off between VTE and natural variation in chorismate metabolism explained by transcriptional reprogramming of specific structural genes of the pathway. By restoring the accumulation of α-t3 (alfa tocotrienols) in fruits, the plants produced here are of high pharmacological and nutritional interest.

Introduction

Tocopherols and tocotrienols, collectively known as tocochromanols, are amphiphilic lipids with VTE activity and as such are essential components of both human and animal diets (Epstein *et al.,* 1966; Evans & Bishop 1922). They are mainly synthesized by photosynthetic organisms (Sussmann *et al.,* 2017) through the condensation of homogentisate (HGA), derived from the shikimate (SK) pathway, with an isoprene chain, which arises either from phytyldiphosphate in the case of tocopherols or from geranylgeranyldiphosphate in the case of tocotrienols . Both prenyl tails can come from the methyl erythritol phosphate (MEP) pathway, alternatively phytyldiphosphate can originate from chlorophyll recycling. Four major forms of tocopherols and tocotrienols occur in plants (α, β, γ and δ), differing in their position and number of methyl groups. Alpha-tocopherol (α-t) is mostly accumulated in leaves whereas γ-tocopherol (γ-t) is the major species found in seeds (Munne-Bosch & Alegre, 2002).

The production of these molecules is a metabolic innovation allowing plants to live under high-oxygen conditions (Zeng *et al.,* 2014) as their role protecting PSII against oxygen singlets and limiting lipid peroxidation are well documented (Kreigeuer-Liszkay & Trebst, 2006; Miret & Munne-Bosch, 2015). More recently, it was demonstrated that the tocopherol biosynthetic pathway modulates salicylic acid accumulation affecting basal resistance against *Pseudomonas syringae* in Arabidopsis plants (Stahl *et al.,* 2019). Furthermore, tocopherols were found to be required for a retrograde chloroplast-to-nucleus signaling mechanism favoring miRNA biogenesis under abiotic stress (Fang *et al.,* 2019). Particularly in tomato (*Solanum lycopersicum*) tocopherol has been demonstrated to be essential in tolerance to combined light and temperature stress (Spicher *et al.,* 2017).

An essential function of tocochromanols in all living cells, irrespective of their precise mechanism of action, is to protect biological membranes against oxidative stresses. In mammals, α-t and α-tocotrienol (α-t3) have the highest antioxidant activities (Suarna *et al.,* 1993). However, some properties unique to tocotrienols have been recently recognized (Peh *et al.,* 2016; Zainal *et al*., 2019), rendering this molecule of particular pharmacological interest. For these reasons, there is a considerable interest in crop plant species with increased or customized tocochromanol contents (Mène-Saffrané & Pellaud, 2017). Attempts to reach this goal were achieved by overexpressing the

structural genes involved in the biosynthesis of tocopherols in several plant species. Increases in tocochromanol accumulation by the overexpression of the cloned pathway gene (encoding γ-tocopherol methyl transferase (VTE4)), have been reported in *Arabidopsis* (Shintani & DellaPenna, 1998), soybean (Van Eenennaam *et al.,* 2003; Arun *et al.,* 2014), shiso (Lee *et al.,* 2008), lettuce (Lee *et al.,* 2005), mustard (Yusuf & Sarin, 2007), maize (Zhang *et al.,* 2013) and tobacco (Jin & Daniell, 2014). Other metabolic engineering approaches focused on the availability of HGA as a substrate for tocopherol, plastiquinone and tocotrienol biosynthesis catalyzed by homogentisate phytyltransferase (HPT), homogentisate solanesyl transferase (HST) and homogentisate geranylgeranyl transferase (HGGT) (Mène-Saffrané & Pellaud, 2017). These approaches were a partial success being somewhat constrained by the limited availability of phytyl pyrophosphate which mostly comes from the degradation and recycling of chlorophylls (Chander *et al.,* 2008; Zhang *et al.,* 2014; *Dorp et al.,* 2015; Almeida *et al.,* 2016). Genetic studies performed in tomato revealed the existence of four metabolic QTL (mQTL) affecting total tocopherol levels in fruits (Almeida *et al.,* 2011). Furthermore, several candidate genes found in this study were functionally characterized and demonstrated to affect this trait through different regulatory mechanisms (Quadrana *et al.,* 2011; Almeida *et al.,* 2016; Bermudez *et al.,* 2018).

Up to now, the natural intraspecific variability of tocochromanols has rarely been explored in tomato (Tieman *et al.,* 2017). Instead, the above mentioned QTL mapping studies focused on introgression lines between cultivated tomato and its wild relatives, *S. chmielewskii* and *S. pennellii* (Ballester *et al.,* 2016; Barrantes *et al.,* 2016; Do *et al.,* 2010; Eshed & Zamir, 1996). One limitation of using such biparental mapping populations is that the limited number of alleles tested are not necessarily representative of the crop natural variation (Blanca *et al.,* 2015; Ranc *et al.,* 2008). Complementary experimental approaches based on multiparental mapping populations and panels assembled for Genome Wide Association Studies (GWAS) attempt to overcome these limitations by incorporating a larger proportion of the extant genetic variation. Recent reports demonstrated the success of such methodologies to identifying genes involved in tocochromanol contents in corn kernels *(*Diepenbrock *et al.,* 2017; Wang *et al.,* 2018).

In the present study, we applied a combination of interval and GWAS mapping in order to investigate the genetic architecture of tocochromanol content in tomato fruits. We

analyzed fruit tocochromanol content in 124 lines of a tomato Multi Parental Advanced Generation Intercross (MAGIC) population grown under control (C) condition (Pascual *et al.,* 2015) and 136 highly diverse small fruit tomato accessions grown under control (C) and water deficit (D) conditions (Albert *et al.,* 2016). These populations were genotyped with 7,400 SNP markers and a total of 25 QTLs were mapped in regions enriched in genes of the amino acid metabolism category. Ten of these genes encode enzymes of the shikimate (SK) and post-chorismate branch pathways. Indeed, the effect of by passing this highly regulated point by the expression of a yeast prephenate dehydrogenase (PDH), which catalyzes the transformation of prephenate into hydroxyphenylpyruvate (HPP), in tomato fruits overexpressing an Arabidopsis hydroxyphenylpyruvate dehydrogenase (HPPD), resulted in massive accumulations of tocotrienols and to a lesser extent tocopherols, at the expense of significant reductions in the levels of tyrosine, phenylalanine and tryptophan during the expansion and ripening phases of fruit development. These results are discussed in the context of the inter-regulation of vitamin E and aromatic amino acid contents and more broadly with respect to the co-ordinate control of plant antioxidant metabolism.

Results

Trait architecture in the MAGIC- and GWAS-mapping populations

The two populations used in this work were studies to evaluate the intraspecific variation of tocochromanol content in tomato fruit. They constitute to a MAGIC population derived from the intercross between four cherry and four large fruited tomato accessions (Pascual *et al.,* 2015) and a GWAS panel of cherry tomato accessions chosen for representing a maximum of the diversity in *S. lycopersicum* cv *cerasiforme* (Albert *et al.,* 2016) and challenged under two irrigation conditions: control (C) and deficient (D). Tocochromanol (α-t, β-t, γ-t, δ-t, α-t3 and tot-t) contents in ripe fruits from both populations displayed large variations (Figure. 1a,b and Figure 1Sa,b) ,with variance analysis revealing high and significant genotype effects (*P*-values < 0.001) representing between 33.2 (δ-t) and 67.6% (α-t3) of the total sum of square. Watering regime effects were small but significant for most of the traits (with the exception of α-t, β-t and tot-t which not presented significant differences) (Figure 1b). GxE interactions were detected for α-t, γ-t, α-t3 and tot-t (Figure 1b). Water availability had a genotype-dependent effect on the fruit tocochromanols contents. In the GWAS panel, distributions from plants cultivated under water deficit overlap with those from plants cultivated under irrigation conditions (Figure 1Sb). While a group of accessions exhibited higher levels of tocochromanols forms under normal irrigation conditions, for a few other accessions displayed clearly lower levels under these conditions (Figure 1a). Heritability values estimated from the GWAS data for each phenotype ranged between 0.234 (for tot-t under D) and 0.612 (for α-t3 under D). These values are comparable with values previously reported for other primary metabolites in tomato (Sauvage *et al.,* 2014; Schauer *et al.,* 2008). For all phenotypes, heritability values were highly similar between the two environmental conditions (Table 1). Interestingly, heritability values estimated from the MAGIC population data were markedly lower and ranged from 0.08 to 0.20 for α-t3 and δ-t, respectively.

Interval mapping and GWAS of tocochromanol contents in tomato fruit revealed an enrichment of amino acid metabolism associated genes amongst the candidate genes

In a first attempt to unveil the genetic bases of these traits we applied interval mapping (IM) to genotypic (~1,300 SNPs) and phenotypic data of 124 RILs from the MAGIC population cultivated under C condition. This approach resulted in the identification of seven QTL (3 for α-t, 2 for δ-t and 2 for tot-t) which were mapped onto five genomic regions on chromosomes 3, 7, 8, 11 and 12 (Figure 2 and Table S1). Percentages of the explained phenotypic variance (PVE) ranged between 12.67 and 18.03 for QTL07.1 and QTL08.1 (α-t and tot-t), respectively. By filtering out the polymorphisms in parental lines according to the expected allelic effects (Pascual *et al.,* 2015), the number of genes within these regions varied between 52 and 208 for the α-t QTL11.1 and the α-t/tot-t QTL08.1, respectively. These QTL, along with previous results reported by our group (Almeida *et al.,* 2011; Quadrana *et al.,* 2014), demonstrate the polygenic architecture of tocochromanols in tomato fruits. Thus, we incorporated a larger amount of genetic variation in order to perform association mapping by applying a multi-locus model (Segura *et al.,* 2012). Through the phenotypic evaluation of a GWAS panel (composed of 136 highly diverse cultivated accessions; Albert *et al.,* 2016) cultivated under two water regime conditions, 18 significant ($p < 10^{-4}$) associations were detected, with ten and eight associations being found under C and D conditions, respectively (Figure S2, Table S2). These associations were mapped onto 16 loci were distributed across nine of the 12 tomato chromosomes (Figure 2 and Table S2). Two genetic markers were significantly associated to two different tocochromanol forms; S02_46637841 (chromosome 2) with α-t and tot-t and S06_42560313 (chromosome 6) with α-t3 and δ-t (Figure 2 and Figure S2). Genomic regions spanning the identified QTL ranged from 30 to 1,830 Kbp harboring between three and 208 annotated genes. The percentage of associated variation of each tococromanol forms (PVE) were relatively variable for each individual association; from 10.4 to 19.4 % (Table 1). These PVE values were comparable with those estimated for the QTL detected with the MAGIC population (Table S4).

Enrichment analyses of all genes in the QTL confidence intervals (3,296 and 958 from the IM analyses and GWAS, respectively) using Mapman categories allowed the identification of eight over-represented categories (against the compete genome, Figure 2). The top three ranking significantly overrepresented categories ($p<0.0001$, Fisher test) were: tetrapyrrole synthesis (+1.60), amino acid metabolism (+1.26) and mitochondrial

electron transport (+12.5). We thus concentrated on the list of annotated loci within these genomic regions with the aim to *a priori* identify candidate genes. The criteria followed to choose these candidates were: (i) that they encode structural genes (or transport proteins) of tocochromanol-related pathways, (ii) that they were previously identified as tocochromanol-candidates genes in tomato and other plant species (Almeida *et al.,* 2011; Diepenbrock *et al.,* 2017),(iii) that they encode structural genes of lipid metabolism pathways or (iv) they are expressed in fruit tissues (data available in Fernandez-Pozo *et al.,* 2017; Zouine *et al.,* 2017). Overall, these analyses resulted in the identification of the 70 candidate genes listed in Table S3. Their chromosomal localizations are shown in the map of Figure 2.

Tocochromanol associated genomic regions include a priori candidate genes enriched in chorismate–tyrosine metabolic pathways

Five of the identified *loci* were in linkage disequilibrium with structural genes of the chorismate–tyrosine metabolic pathways (Figure 3a and Table S3). Moreover, IM results pinpointed two QTL harboring five other structural genes of these same pathways (Figure 3c). Overall, the six QTL harboring these 10 candidate genes explained between 33.3 and 75.2 % (for α-t3 and tot-t, respectively) of the PVE by the total identified QTL (Table 1 and Table S3). Significant allelic effects within the GWAS population and founder effects for the different parental alleles of the MAGIC population were also detected for these loci (Figure 3b,d). Additionally, subsequent haplotype analyses on the above *a priori* candidate genes revealed significant allelic effects for four of them; *TyRA2* (Solyc09g011870), *fbt* (Solyc06g068560), *ADCL* (Solyc11g071280) and *DHQS* (Solyc02g083590) (Figure S3). These genes encode regulatory enzymes of the shikimate pathway and catalyze reactions from 3-deoxy-D-arabino-heptulosonate-7-phosphate and homogentisate, the supplier of the aromatic head group of the tocochromanol molecules (Figure 4a). Reactions from chorismate to tyrosine conversions have been previously reported as being highly sensitive to end-product regulation (Tzin & Galili 2010). Given that this mechanism involves bottleneck steps of regulation, we engineered tomato plants by-passing these reactions in order to functionally validate our genetic analyses.

Engineering chorismate-tyrosine pathway in tomato fruits

We concentrated on engineering the chorismate-tyrosine pathway in tomato fruits using a transgenic approach wherein the TyrA-catalyzed reaction is bypassed by the fruit specific expression of a prephenate dehydrogenase encoding gene from yeast (*ScPDH*). The subsequent reaction of *p*-hydroxyphenylpyruvate conversion into homogentisate was boosted by the expression of an *Arabidopsis* encoding hydroxyphenylpiruvate dioxigenase gene (*AtHPPD*) (Figure 4a; Figure S4a). Both genes were expressed under control of PPC2 promoter of the phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31) gene from tomato. After the first round of Agrobacterium transformation, 85 tomato primary transformants (T0) were obtained. Among all these transgenic lines grown in greenhouse, none of them displayed phenotypic changes at the plant developmental and fruit (growth, size, color) levels. Molecular and compositional analyses of these primary transformants (T0) revealed detectable expression of both transgenes in fruits from five of these transgenic lines (Figure S4b). Fruits at breaker+4 days stage (orange) from three of these T0 lines (PH13, PH17 and PH30) showed tocochromanol compositional changes (Figure 4b) and/or contents (Figure S4c) with respect to the wild type controls. All three lines showed enrichments in total tocotrienols (tot-t3) content and PH13 line also showed ~25% more tot-t than wild-type fruits.

Chorismate–tyrosine pathway engineered plants accumulate more tocochromanols

In order to confirm the results presented above in a subsequent generation (T1) we analyzed the expression of the transgenes (by qPCR) along with tocochromanol composition at three developmental stages of the fruits (5 days post anthesis, breaker and breaker+8 days) from five double-homozygous lines (Figure 5). While the expression levels of both transgenes were substantially high in fruits from the three PH13-sibling lines, PH17 line showed higher *ScPDH* transcript levels than those from *AtHPPD* gene. The opposite was true for the case of PH30 line (Figure 5a). These results were together with the tocochromanol levels observed in these same fruits; marginal increments (~20%) in tot-t contents and massive accumulations of tot-t3 observed in the PH13-sibling lines and unchanged levels, similar to the WT for PH17 and PH30 lines (Figure 5b).

Robustness of these results were confirmed in a parallel experiment where a wider range of fruit developmental stages were analyzed for the PH13 sibling lines (from 5 days

post anthesis (5 DPA) to breaker plus 8 days (Br+8)) (Figure S5a-f). Increases in the levels of α- and β-t were observed. By contrast, γ-t levels decreased significantly in the fruits of the transgenics at 30 DPA and at later stages (breaker and ripe) (Figure S5b and Table S4). Intriguingly, both α-t and γ-t showed opposite profiles of the t3 forms along development and ripening (Figure S5b). These results suggest that the observed increments in tocotrienol levels are not only due to the expected increments in homogentisate pool but also occur at expenses of flux into the tocopherols (Figure S5a,b). In addition α+β/γ+δ tocopherol ratio, which is indicative of the efficiency of the last step of the pathway (mediated by the VTE4 enzyme) displayed a clear decrease, markedly at cell division stages (5 and 10 DPA) (Figure S5c). The main cause of this gradient is the observed decay in α-t contents and an opposite increase of the γ-t form in both PH13 and WT plants (Figure S5d). In accordance with the previous observation, the α+β/γ+δ -t3 ratio showed the opposite behavior (Figure S5e,f), which also suggests that the observed accumulation of t3 occurs at expenses of tocopherols.

As expected, the tocochromanol contents in source leaves of these transgenic plants were invariant from wild-type controls (Figure S6). This is in agreement with the fact that the *SlPPC2* promoter is only functional in fruits.

Along with these variations in the tocochromanol levels, pigments such as chlorophyll, carotene and lycopene presented significant differences at specific fruit stages without clear accumulation trend in the transgenic lines (Table S5). In accordance with the observed increases in tocochromanol levels, electron microscopy images of green fruits demonstrated a significant increase in the size and number of plastoglobules per chloroplast in the three PH13-deribed sibling lines (Figure 6a,b,c). These changes resulted in a remarkable increase in the ratio of plastoglobule/plastid per chloroplast. However, these results should be taken with caution as they were analyzed in three sibling lines derived from a single transgenic event (PH13); and an accurate validation in other independent lines should be necessary. Notwithstanding these visual changes in chloroplast ultrastructure, the main lipids detected in tomato fruits (FAPEs, TAG, MGDG, DGDG and PL) were invariant between WT and sibling PH13 lines (Table S6).

The increased tocochromanol content impacts aromatic amino acid metabolism

In order to better understand how the engineered plants were affected at the transcript level, a dedicated qPCR array including tocochromanol biosynthetic genes (Quadrana *et al.,* 2013) was applied to fruits at three different stages of development (20 DPA, BR and BR+8). A PCA biplot based on the mRNA levels of the 23 genes analyzed showed that 62.4% of the variation is comprised in the first two components. The different lines grouped firstly according to fruit maturity and secondly by genotype evidencing a different mRNA profile of the double transgenic line PH13 in comparison with PH17, PH30 and WT (Figure 7a). By inspecting the loadings it is possible to identify those genes whose expression levels best define the different developmental stages.Most genes were down-regulated in fruits of the double-expressing line PH13 particularly at BR and BR+8 stages (Figure 7b). Indeed exceptions to this behavior were only detected for the by-pass enzyme encoding genes; arogenate dehydrogenase (*TyrA(1)* and *TyrA(2)*), tyrosine aminotransferase (*TAT(2)*) and hydroxyphenylpyruvate dioxygenase (*HPPD(2)*), whose transcripts increased in the fruits of the double transgenic line. A further exception is *HST* which displays a differential expression pattern with a peak at BR+8. Alongside these changes in transcripts, the fruits of PH13 line showed inverse correlations between tocochromanols and aromatic amino acid profiles, which is in agreement with the expression of the biosynthetic genes detected in these transgenics (Figure 5a and Figure 7b), highlighting a metabolic tradeoff between these components during development.

Discussion

With the aim of elucidating the genetic architecture of the complex trait of VTE content in tomato alongside defining the regulatory mechanisms which determine the flux through it biosynthetic pathway, we evaluated the extent of tocochormanol natural variation in fruits of 124 F7 lines of a MAGIC population (Albert *et al.,* 2016) and 136 accessions of a tomato diversity panel (Pascual *et al.,* 2015). Exploration of variation in tocochromanol levels in these two populations revealed a vast range in the contents of the five forms evaluated here. Whereas the MAGIC population was evaluated only under control condition, the GWAS population was grown under control and water deficit conditions partitioning the variation observed into genotype, watering regime and genotype by watering regime effects. The phenotypic variations observed for tocochromanol contents in the diversity panel resulted mainly from large genotype effects (33.2–67.6%) and medium to low watering regime effect (0.2-3.0%) and genotype by watering regime interactions (9.4-17.3%). As a consequence, moderate to high heritabilities were observed for these traits, similarly to the range of heritability previously reported in tomato (Sauvage *et al.,* 2014).

Combining two complementary mapping approaches, IM and GWAS, we identified a total of 25 QTL. Indeed, VTE content has been described as a multilocus trait not only in tomato (Almeida *et al.,* 2011; Quadrana *et al.,* 2013) but also in several other plant species (Arabidopsis -Gilliland *et al.,* 2006-; maize -Diepenbrock *et al.,* 2017-; rape seeds –Wang *et al.,* 2012-; rice -Sookwong *et al.,* 2009- and soybean -Li *et al.,* 2016-). Most of the QTL previously identified using 9 ILs of *S. pennellii* (Almeida *et al*., 2011) map on chromosomes 7, 8, and 9 and overlap with 4 of the ones detected here. This could be a consequence of the high genetic divergence between *S. lycopersicum* and *S. pennellii*.

The detected amount of α-t3 in fruits from both mapping populations is striking as modern tomato varieties do not accumulate this molecule (Horvath *et al.,* 2006; Chun *et al.,* 2006; Quadrana *et al.,* 2013). Large variation in α-t3 content could be explained by the origin and composition of the genetic material used in this study; mainly cherry tomatoes from South America and heirloom varieties, which permitted us to map seven QTL for this tocochromanol form. More importantly, the strategy to functionally validate these genetic analyses resulted in engineering tomato plants with high pharmacological and nutritional values due to the restored capability to massively accumulate α-t3 in fruits.

This form of VTE has been shown to possess unique properties such as the inhibition of 3-hydroxy-3-methylglutaryl-coenzyme A reductase which leads to a lowering of cholesterol levels, attenuation of inflammation via downregulation of transcription factor NF-κB activation, and acts as a potent protectant against radiation damage (reviewed by Peh *et al.,* 2016). In addition eighteen other associations were detected for α-t, δ-t and tot-t. Two associations (S02_46637841 and S06_42560313) were in common for two different tocochromanol forms (α-t/ tot-t and α-t3/ δ-t) as can be expected for metabolites that share structural genes in their biosynthetic pathway.

As previously mentioned, tocochromanol content in tomato fruit is defined as a polygenic trait. This is in agreement with both the number of detected QTL (25) as well as the percentages of variation explained by each of them, which ranged between 10.4 and 18.1 %. Nevertheless, the 10 candidate genes encoding structural genes of the shikimate pathway were identified in 6 of the 25 QTL whose PVE showed major effects (between 33.3 and 75.2 % for the cases of α-t3 and tot-t, respectively). One of such candidate loci, encoding the TyRA(2) enzyme maps to a previously reported a QTL for α-, β-, γ- and tot-t contents in tomato fruits (Almeida *et al.,* 2011). Haplotype analyses of this locus (Solyc09g011870) revealed significant allelic effects and high divergence between 45 accessions of the 136 included in the GWAS panel. This could cause differences in the enzymatic activity and a high flux of hydroxyphenylpyruvate to the central route of tocochromanols biosynthesis. It has been previously demonstrated that altering *TyRA* and *CM* expression levels in Arabidopsis has an impact in the accumulation of tocochromanols both in seeds and leaves (Tzin *et al.,* 2009, Zhang *et al.,* 2013; Cahoon *et al.,* 2012; de Oliveira *et al.,* 2018). Moreover, *TyrA* overexpression in soybean results in massive accumulations of tocochromanols (Valentin *et al.,* 2005). Similar phenotypes were observed also in soybeans plants lacking the HGO enzyme which catalyzes the degradation of HGA (Stacy *et al.,* 2016). At the same time, this gene was identified as a candidate in a GWAS for tocochromanol contents in maize kernels (Diepenbrock *et al.,* 2017).

Others two candidate genes showing significant allelic effects on α-t3 and α-t are *fbt* and *ADCL*, respectively, whose products divert chorismate to folate biosynthesis thereby representing a trade-off between VTE and vitamin B9. On theother hand, *AnPRT*, another of the candidates identified here (two paralogues *(1)/(2)*) catalyzes one of the

first committed steps in Trp biosynthesis and as such also requires chorismate. A paralogue of these loci have been found also associated to a β-t QTL (Almeida *et al.,* 2011). Tyrosine can also be exported from plastid to cytoplasm were HGA synthesis is localized (Muñoz & Munné-Bosch, 2019) by the action of *LAT1*, an ortholog tyrosine transporter of araport11 (At2g33260), ultimately affecting tocochromanol biosynthesis. This subset of candidate genes encode enzymes and transporters of the chorismate-tyrosine pathway, a highly regulated branching point (Tzin & Galili, 2010). Taken alongside the results from the enrichment analyses which revealed that amino acids metabolism was a highly overrepresented category, led us to engineering tomato plants by-passing these reactions in order to functionally validate our genetic analyses. With this aim, we produced a set of transgenic lines co-overexpressing the *ScPDH* and the *AtHPPD* genes in order to boost HPP and HGA fluxes in a fruit specific manner (Fernandez *et al.,* 2009). Detailed characterization of these lines showed significant increases of tocotrienols (up to ~ 3,400%) and moderate increases in tocopherols (up to ~ 20%).

These increases in tocochromanols in the fruits of the PH13 line observed at both T0 and T1 generations are indicative of a minimum expression threshold of both genes as have been suggested by Rippert *et al.* (2004) for leaf tissue of tobacco using these same genes under the control of a constitutive promoter. Neither fruits from PH17 line, with high expression of *scPDH* gene nor those from PH30 line with high levels of *AtHPPD* mRNA were enough to achieve the levels of tocochromanols showed by fruits of the PH13-sibling lines.

It is worth mentioning to remark that the maximum accumulation of tocochromanols occurs at the BR stage of the fruits in accordance with the peak of *PPC2* promoter activity (Guillet et al., 2012; Fernandez et al., 2009; Nafati et al., 2011). This suggest that such increases in tocochromanols is the result of high HGA synthesis and enough availability of the other substrate geranylgeranyl-DP.

The enzyme homogentisate geranylgeranyl-DP transferase (HGGT) was not yet identified in tomato. Tocotrienol increases could, however, be explained by the promiscuous activity in substrate acceptance of other prenyl transferases such as homogentisate solanesyl transferase (HST) and/or the homogentisate phytyl transferase (VTE2) (Almeida *et al.,* 2011; Sadre *et al.,* 2006; Tian *et al.,* 2007) providing further

evidence in support of tocotrienol production in the absence of a specific HGGT. That said, other reports demonstrate increased tocopherols in the absence of tocotrienols in transgenic *Arabidopsis* and tomato plants over-expressing *VTE2* (Collakova & DellaPenna 2003; Lu *et al.,* 2013). These findings explain why tocopherol synthesis is not limited by the availability of HGA and PDP precursors but rather depend on the activity of the *VTE2* enzyme. Nevertheless, tocotrienols are limited by HGA levels in tomato fruits.

As shown for the majority of the 23 analyzed genes, VTE2 expression is down-regulated in fruits of the PH13 lines, particularly at BR and BR+8 stages. By contrast, the *HST* gene as well as those encoding the by-passed enzyme (TyrA(1)/(2), TAT(2) and HPPD(2)) are up-regulated in the fruits of the double transgenic line. Along with these transcriptional changes, the amino acids Tyr, Phe and Trp decreased in the fruits of the PH13-sibling lines, indicating a transcriptional reprogramming of aromatic amino acid metabolism and highlighting a sophisticated regulatory circuitry that the plant cell employs to balance tocochromanols contents.

In summary, the work presented here demonstrates the power of combining different mapping strategies (GWAS and MAGIC panels) to characterize the genetic architecture of complex traits. Functional validation of these results resulted in the production of tomato fruits with enhanced VTE contents. Massive increases in tocotrienol levels could be explained, at least in part, by a transcriptional impairment of specific genes from the chorismate-tyrosine pathway which directly impacted aromatic amino amino acid contents. Indeed, tocotrienols have been shown to exceed VTE activity of tocopherols in a number of model membrane studies (Packer *et al.,* 2001; Peh *et al.,* 2016; Serbinova & Packer 1994; Suzuki *et al.,* 1993) and thus, the plants produced here constitute a high value nutritional biotechnological product.

Experimental procedures

Plant materials and growth conditions

Two populations were used in this study: (i) A subset of 124 lines of the MAGIC population previously obtained by crossing eight tomato, selected to include a wide range of genetic diversity (Table S7) (Pascual *et al.,* 2015); (ii) The GWAS diversity panel consisted of 136 accessions of small fruit tomato. These accessions were previously described in Albert *et al.* (2016) and Blanca *et al.* (2015). The genetic structure of the GWAS panel reflected the species and the geographic origin of the accessions (Table S8). Ten accessions were *S. pimpinellifolium* (SP; the closest wild ancestor of the tomato) originating from Peru and Ecuador. A total of 107 accessions were *S. lycopersicum* var. *cerasiforme* (SLC) originating mainly from South America. Finally, 19 accessions belonged to a mixed genetic group mainly including commercial cherry tomatoes and admixed genotypes between SP, SLC, and *S. lycopersicum* var. *lycopersicum*.

Experimental design

The plants were cultivated with the same experimental design as in Albert *et al.* (2016). Plants were grown in a heated glasshouse in INRAE Avignon, France from March to July 2014. Two watering regimes were applied to the plants: control (C) and water deficit (D). The control treatment was set according to evapotranspiration coefficient for tomato under greenhouse conditions (FAO Water, 2015). A maximal drainage of 25% and a relative humidity of the substrate of 65% were established in the control pots. Water deficit treatment was applied progressively after the flowering of the second truss of the earliest accession. Watering was first reduced by 25% compared with the control for 1 week and then reduced by 60% until the end of the experiments. Relative humidity of the peat substrate was controlled with GRODAN® moisture probes and monitored between 25% and 30% in drought pots. Three plants per watering regimes per accession of GWAS panel were randomized in the greenhouse. One replicate per line of each of the 124 MAGIC was grown at control watering regime, including two F1 replicates and two parental lines.

Tocochromanol phenotyping

Tocopherol and tocotrienol extractions and quantifications were performed as described by Fraser *et al.* (2000). Averaged non-transformed data from 2-3 replicates of the GWAS accession panel are available in Table S8. Table S7 contains data from the 124 MAGIC lines, the F1 and parentals.

Genotyping and SNP filtering of the GWAS panel

The GWAS population was genotyped using the Tomato Infinium Array developed within the SolCAP project (Hamilton *et al.,* 2012; Sim *et al.,* 2012). After filtering, the set of markers consisted of 6100 SNPs see Albert *et al.* (2016).The SNPs were renamed according to their positions on the tomato genome for example SL2.50 was renamed as S01_58000085 given that it represented base pair 58000085 on chromosome 1. QTL mapping in the MAGIC population was performed with a set of 1300 SNP markers as described in Pascual *et al.* (2015).

Statistical analysis of the phenotypic data for GWAS

Phenotypic data from both conditions were analyzed separately (R Core Team., 2012). They were normalized using Box and Cox transformations. Then the ANOVAs were performed according to the following model:

$Y_{ijkl} = \mu + G_i + W_k + G_i \times W_k + e_{ijkl}$

Y_{ijkl} was the phenotypic value of accession i in watering regime k, μ the overall mean, G_i the fixed genetic effect, W_k the fixed effect of watering regime k, $G_i \times W_k$ the fixed effect of interaction of both and e_{ijkl} the residual error effect. Phenotypic variation measured in the MAGIC population showed normal distribution after Box-Cox transformation (Figure S1a). Similarly, transformed data from the GWAS population showed a normal distribution in the cases of α-t, γ-t and total-tocopherols (tot-t) but bi-modal distributions for the cases of α-tocotrienol (α-t3) and β-tocopherol (β-t) (Figure S1b). However, neither a particular genetic group or subgroup, nor the color of the fruits were associated to these bi-modal distributions.

Best linear unbiased predictions (BLUPs) were estimated using the LME4 package in R (Bates *et al.,* 2007). Several models were evaluated with the best model chosen

based on the AIC criteria. Finally a model adapted from Baseggio *et al.* (2018) was used as follows:

$Y_{ijkl} = \mu + G_i + M_i + W_k + G_i x\ W_k + e_{ijkl}$

Interval mapping in the MAGIC population

For interval mapping, Box Cox transformation of the phenotypic raw data was performed, subsequently the mpMap package (Huang & George 2011) was used based on the founder's parental probabilities as input for QTL identification in the MAGIC population. Previously, the mpMap function computes founder effects at a step size of 2 cM with a regression approach, based on the multipoint probabilities computed with 'mpprob' function. QTL were called when *p* -values were smaller than the empirical threshold *p*-value (1.72×10^{-4}) derived using the function 'sim.sigthr' after computing 1000 permutations, to reflect a genome-wide significance threshold of 0.05. QTL support intervals were determined with a 1-LOD drop support. After QTL detection, the function 'fit' from R mpMap package to estimate the percentage of phenotypic variation explained by the QTLs. Then, all the polymorphisms (Causse *et al.,* 2013) present in the QTL support intervals were analyzed and filtered according to the estimated founder effects. Heritability values were estimated for all phenotypic traits following the method proposed by Broman *et al.* (2019).

GWA mapping

Best linear unbiased predictions (BLUPs) were used in the mapping models. GWAS were performed using correction for population structure (PCoA) and modeling genetic variance with the kinship matrix (K) based on identity by state among the 6100 SNPs. We used the multi-locus mixed-model (MLMM) developed by Segura *et al.* (2012) to increase the detection power for polygenic characters. Models with a maximum of five cofactors all having a raw P-value $<10^{-4}$ were retained. From the optimal model selected, the percentage variation explained by the selected markers (global PVE for all the significant markers) was computed for each tocochromanol form.

Linkage disequilibrium

Linkage disequilibrium (LD) between markers on each chromosome was calculated using the R2 estimator implemented in the package 'genetics' (Warnes & Leisch, 2011). First, we performed LD calculation between 100,000 randomly chosen pairs of unlinked loci (on different chromosomes). The 95th percentile of the unlinked-R2 distribution equal to 0.28 was considered as the critical LD threshold. Then, for each significant marker, we computed LD with all the markers upstream and downstream on the same chromosome. Through plotting of R2 versus genetic distance between markers we defined both boundaries of the confident interval as the intersection of the 'critical LD' threshold and the fitted curve of R2 regression (Figure S7). A nonlinear model described by Remington & Thornsberry (2001) was used to fit the LD decay. The number of genes within each interval was identified from the tomato genome (ITAG2.4).

Haplotype analyses of candidate genes detected by GWAS.

Based on the nonreference genome sequences of the tomato pan-genome (Gao *et al.,* 2019) we defined haplotype blocks for five of the 70 candidate genes belonging to the chorismate-tyrosine pathway (Figure S3). We extracted the sequences for those five *loci* for 45-50 of the 136 accessions, if more than 50% of the CDS region was covered by contigs. Then we performed multiple nucleotide sequence alignments, 2 Kpb up and downstream of each gene using MUSCLE (Edgar, 2004). This subset of accessions is representative of the total 136 accessions, considering previous analysis for genetic structure all groups and subgroups were represented (Albert *et al.,* 2016 and Table S2). The frequencies for each haplotypes were calculated as a percentage over those 45-50 accessions. The Kruskal Wallis test was used for estimation of phenotypic differences between haplotypes, considering each biological replicate. Only non-synonymous substitutions were chosen to define haplotype blocks but hundreds of polymorphisms were found 2 Kpb up and downstream of each analyzed loci. Prediction for non-synonymous amino acid substitutions was evaluated using SIFT software in order to evaluate functional effects (Ng & Henikoff, 2003).

Enrichment analysis by Mapman categories

Ease score, a modified Fisher Exact *P*-value, was used for gene-enrichment analysis. Ease score <0.0001 was considered strongly enriched in the annotation

categories. The functional enrichment analyses of typical Mapman gene categories were based on the functional annotated file Slyc_ITAG2.4 (https://MapMan.gabipd.org/MapManstore).

Trangenics plants material and growth conditions

Cherry tomato [*Solanum lycopersicum* L. cv West Virginia 106 (WVa106)] plants were grown in a greenhouse or *in vitro* as described by Alhagdow *et al.* (2007). Several plants over-expressing in the fruits both the *ScPDH* and *AtHPPD* genes under the control of the *SlPPC2* promoter, named as *pdh/Athppd*^{OE} lines or more simply PH lines, were selected. Primary transformants were checked for ploidy level by flow cytometry. In diploid plants, the presence of *ScPDH* and *AtHPPD* constructs was checked by PCR using the specific primers and only the lines displaying the two constructs were considered for this study. Tomato fruits were harvested at various developmental stages according to the number of days post-anthesis (DPA) and fruit diameter. Prior to all biochemical and molecular analyses, the jelly part and the seeds were removed from the fruits and samples were quickly frozen in liquid nitrogen, ground to fine powders and stored at -80°C until use.

Cloning procedures

The sequences encoding the OTP-*ScPDH* (*Tyr1*, accession no. Z36035) and OTP-*AtHPPD* (HPPD, AT1G06570-1) were amplified by PCR using the respective plasmids pPCR Script-OTP-PDH and pRPA 150 A2-ARA9 (Garcia *et al.,* 1999; Rippert *et al.,* 2004) and were then inserted into the GATEWAY vector pDONR201 (Invitrogen, http://www.invitrogen.com/) by attB recombination following the manufacturer's protocol. The second recombination reaction was performed using the destination pK2GW7-pPPC2 vector containing a 2-kb DNA fragment of the tomato phosphoenolpyruvate carboxylase-2-promoter (*SlPPC2*), a cell expansion phase fruit-specific promoter (Fernandez *et al.,* 2009). The two vectors, pK2GW7-p*Sl*PPC2-OTP-*At*HPPD and pK2GW7- p*Sl*PPC2-OTP-*Sc*PDH, were separately inserted into *Agrobacterium tumefaciens* strain GV3101 by transformation. A mixture constituted of half of each of the two agrobacterium cultures was subsequently used to transform *Solanum lycopersicum*

cv. WVA106 cotyledons and plants were regenerated as described by Alhagdow *et al.* (2007).

RNA Isolation and qPCR analyses

RNA extraction and RT-PCR analyses using gene-specific primers (Table S9) were performed as previously described by Quadrana *et al.* (2013) on several tomato organs, namely leaves, fruits for genes related to the MEP and SK pathways, including also the *ScPDH* and *AtHPPD* genes.

GC-MS metabolite profile analysis

Fruit pericarp prepared as described above were used for metabolite extraction as described by Nunes-Nesi *et al.* (2005). The level of all metabolites was quantified by GC-MS exactly following the protocol described by Roessner *et al.* (2001), with the exception that the peak identification was optimized to tomato tissues (Roessner-Tunali *et al.* 2003). The extraction, derivatization, standard addition, and sample injection for GC-MS were performed according to Lisec *et al.* (2006) and Osorio *et al.* (2012). Identification and quantification of the compounds were performed with TagFinder 4.0 software and the mass spectra were cross-referenced with those in the Golm Metabolome Database (Kopka *et al.,* 2005). Six biological replicates were used for this analysis.

Pigment quantification by HPLC

Chlorophylls, lycopene and carotenoids were extracted and quantified exactly as described in (Quadrana *et al.,* 2013). Lipids were extracted from lyophilized pericarp according to (Folch & Stanley 1987). Fatty acid methyl esters (FAMEs) were then extracted in 400 µL hexane and analyzed by Gas Chromatography performed using an Agilent 7890 gas chromatograph equipped with a Carbowax column (15 m x 0.53 mm, 1.2 µm; Alltech Associates, Deerfield, IL, USA) and flame ionization detection. FAMES were identified by comparing their retention times with commercial fatty acid standards (Sigma-Aldrich) and quantified using ChemStation (Agilent) to calculate the peak surfaces.

Electron microscopy

Chloroplast ultrastructures of 20 DPA WT and PH13 pericarps were examined by electron microscopy. Pericarps were cut in small pieces and fixed in 2.5 % glutaraldehyde in 0.15M phosphate buffer overnight at 4ºC, and post-fixed in 1% osmium tetroxide for 2 h and 1% tannic acid for 1 h at room temperature (22–25ºC). After dehydration and embedding in Epon resin, ultrathin sections (70-80 nm) were obtained on a Reichert Ultracut S microtome and mounted on to copper grids. Ultrathin sections were observed with a FEI Tecnai G2 Spirit TWIN 120kV transmission electronic microscope equipped with a CCD 16M pixels Eagle 4k camera CM 100 transmission electron microscope at 60 kV. Chloroplast area and plastoglobule diameter were measured using the ImageJ software in cell layers 8 and 9 in a centripetally manner from the first epidermis layer.

Principal Component Analysis (PCA)

General overview of the variance for the transcripts levels analyzed in the 3 transgenic lines and WT were represented in a reduced dimensional Euclidean space using BioStatFlow. Data set corresponded to each transgenic plant per line standardized by Z-score transformation as commonly used in PCA analysis.

Data availability statement

All the set of markers reported in this paper are available in Albert *et al.* (2016) and Causse *et al.* (2013).

Acknowledgements

We acknowledge all the experimental teams for their collaboration in implementing the experimentations. We particularly thank Zhangjun Fei for sharing the genome sequences of the tomato pan-genome to define haplotypes block for candidate genes. This work was partially supported by grants from ANPCyT Agencia Nacional de Promoción Científica y Tecnológica, Argentina. (2014-0984 and 2017-1907); European Union Horizon 2020 Research and Innovation Programme (679796); MINCYT-ECOS projects (A11B01 and A14B03).

Author contributions

EB conducted experiments. EB and LdH performed GWAS and ID, QTL mapping of the MAGIC-POP. EB with FC analyzed data and wrote the manuscript. EA and CS developed scripts for the GWAS mapping. EB sampled and collected phenotypic data in France and Argentina. EB, BDL and RA performed the sample collection and tocochromanols measurements. LB and LQ conducted RNA Isolation and qPCR analyses and LB, LQ, ANN and ARF quantified the level of all metabolites by GC-MS. MM, CB, TG, CT, IA, JJ, PB and JPM performed cloning procedures, trangenics plants, pigment quantification by HPLC and Electron microscopy. FC and MC supervised the project, built the experimental design, and revised the manuscript. All authors discussed the results and commented on the manuscript.

Conflict of Interest Statement

The authors declare that they have no competing interests.

Supporting Information

Figure S1. Histogram plots of phenotypes measured in the populations used in this study.

Figure S2. Manhattan and QQplots of the GWAS analysis.

Figure S3. Haplotype analyses of candidate genes detected by GWAS.

Figure S4. Constructs and expression analyses of *ScPDH* and *AtHPPD* transgenic plants at T0 generation.

Figure S5. Accumulation pattern of tocochromanol during tomato fruit development and proportions expressed as α+β/γ+δ in fruits from PH13 lines and wild-type controls.

Figure S6. Levels of tocochromanols (tot-t and tot-t3) in source leaves from PH13 progeny lines and wild-type controls.

Figure S7. Linkage disequilibrium vs distances for each significant association marker of the GWAS associations.

Table S1. QTL identified for tocochromanol contents by interval mapping in the MAGIC population.

Table S2. QTL identified by GWAS for tocochromanol contents in mature tomato fruits under two irrigation conditions.

Table S3. Candidate genes identified within each Interval Confidence.

Table S4. Raw phenotypic data of the tocochromanols contents for mature tomato fruits and leaf tissue from Wildtype and PH13 plants.

Table S5. Analysis of the pigment contents in developing fruits for PH13 and WT.

Table S6. Fatty acid and Lipid composition of 20 DAP fruits from the PH13 and WT.

Table S7. Tocochromanols contents for mature tomato fruits of the 124 MAGIC-POP F7 lines, F1 and Parents.

Table S8. Genetic and phenotypic description of the accessions in the GWA population.

Table S9. Primers information.

References

Albert E, Gricourt J, Causse M, Segura V, Bonnefoi J, Derivot L. 2016. Association mapping reveals the genetic architecture of tomato response to water deficit: focus on major fruit quality traits. J Exp Bot. 67(22):6413–30.

Almeida J, Da Silva Azevedo M, Spicher L, Glauser G, Vom Dorp K, Guyer L, et al. 2016. Down-regulation of tomato PHYTOL KINASE strongly impairs tocopherol biosynthesis and affects prenyllipid metabolism in an organ-specific manner. J Exp Bot. 67(3):919–34.

Almeida J, Quadrana L, Asís R, Setta N, De Godoy F, Bermúdez L, et al. 2011. Genetic dissection of vitamin e biosynthesis in tomato. J Exp Bot. 62(11):3781–98.

Arun M, Subramanyam K, Theboral J, Sivanandhan G, Rajesh M, Kapil Dev G, et al. 2014. Transfer and targeted overexpression of γ-tocopherol Methyltransferase (γ-TMT) gene using seed-specific promoter improves tocopherol composition in Indian soybean cultivars. Appl Biochem Biotechnol. 172(4):1763–76.

Ballester AR, Tikunov Y, Molthoff J, Grandillo S, Viquez-Zamora M, de Vos R, et al. 2016. Identification of Loci Affecting Accumulation of Secondary Metabolites in Tomato Fruit of a Solanum lycopersicum × Solanum chmielewskii Introgression Line Population.Front Plant Sci.7:1428.

Barrantes W, López-Casado G, García-Martínez S, Alonso A, Rubio F, Ruiz JJ, et al. 2016. Exploring New Alleles Involved in Tomato Fruit Quality in an Introgression Line Library of Solanum pimpinellifolium. Front Plant Sci.7:1172. doi: 10.3389/fpls.2016.01172.

Baseggio M, Murray M, Magallanes-lundback M, Kaczmar N, Chamness J, Buckler ES, et al. 2019. Genome-Wide Association and Genomic Prediction Models of Tocochromanols in Fresh Sweet Corn Kernels.1–17.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 67(1),1-48.doi:10.18637/jss.v067.i01.

Bermúdez L, del Pozo T, Lira BS, de Godoy F, Boos I, Romanó C, et al. 2018. A Tomato Tocopherol Binding Protein Sheds Light on Intracellular a-tocopherol Metabolism in Plants. Plant Cell Physiol. 59(11):2188-2203. doi: 10.1093/pcp/pcy191.

Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, et al. 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. BMC Genomics.16:257. doi: 10.1186/s12864-015-1444-1.

Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen S, et al. 2019. R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics. 211(2):495–502.

Chander S, Guo AYQ, Yang AXH, Yan AJB, Zhang AYR, Song ATM, et al. 2008. Genetic dissection of tocopherol content and composition in maize grain using quantitative trait loci analysis and the candidate gene approach. Molecular Breeding. 22:353

Cho EA, Lee CA, Kim YS, Baek SH, de los Reyes BG, Yun SJ. 2005. Expression of gamma-tocopherol methyltransferase transgene improves tocopherol composition in lettuce (Latuca sativa L.).Mol Cells.19(1):16-22.

Collakova E, DellaPenna D. 2003.The Role of Homogentisate Phytyltransferase and Other Tocopherol Pathway Enzymes in the Regulation of Tocopherol Synthesis during Abiotic Stress. Plant Physiol. 133(2):930–40.

de Oliveira MVV, Jin X, Chen X, Griffith D, Batchu S, Maeda HA. 2019. Imbalance of tyrosine by modulating TyrA arogenate dehydrogenases impacts growth and development of Arabidopsis thaliana. Plant J. 97(5):901–22.

Diepenbrock CH, Kandianis CB, Lipka AE, Magallanes-Lundback M, Vaillancourt B, Góngora-Castillo E, et al. 2017. Novel Loci Underlie Natural Variation in Vitamin E Levels in Maize Grain. Plant Cell. 29(10):2374–92.

Do PT, Prudent M, Sulpice R, Causse M, Fernie AR. 2010. The Influence of Fruit Load on the Tomato Pericarp Metabolome in a Solanum chmielewskii Introgression Line Population . Plant Physiol. 154(3): 1128–1142.

Dorp K Vom, Hölzl G, Plohmann C, Eisenhut M, Abraham M, Weber APM, et al. 2015. Remobilization of phytol from chlorophyll degradation is essential for Tocopherol synthesis and growth of arabidopsis. Plant Cell. 27(10):2846–59.

Edgar RC. 2004.MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–7.

Epstein SS, Forsyth J, Saporoschetz IB, Mantel N. 1966. An Exploratory Investigation on the Inhibition of Selected Photosensitizers by Agents of Varying Antioxidant Activity. Radiat Res. 28(2):322.

Eshed Y, Zamir D. 1995. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics . 141(3):1147–62.

Evans & Bishop. 1922. ON THE EXISTENCE OF A HITHERTO UNRECOGNIZED DIETARY FACTOR ESSENTIAL FOR REPRODUCTION. Science. 56(1458):650–651.

Falk J, Munné-Bosch S. 2010. Tocochromanol functions in plants: Antioxidation and beyond. J Exp Bot. 61(6):1549-66. doi: 10.1093/jxb/erq030.

Fang X, Zhao G, Zhang S, Li Y, Gu H, Li Y, et al. 2019. Chloroplast-to-Nucleus Signaling Regulates MicroRNA Biogenesis in Arabidopsis. Dev Cell. 48(3):371-382.e4.

FAO W. Crop WATER INFORMATION.2015.

Fernandez AI, Viron N, Alhagdow M, Karimi M, Jones M, Amsellem Z, et al. 2009. Flexible tools for gene expression and silencing in tomato. Plant Physiol. 151(4):1729–40.

Fernandez-Pozo N, Zheng Y, Snyder SI, Nicolas P, Shinozaki Y, Fei Z, et al. 2017. The Tomato Expression Atlas. Kelso J, editor. Bioinformatics. 33(15):2397–8.

Folch I, Lees M. & Sloane-Stanley GH. 1957. A simple method for the isolation and purification of total lipids from animal tissues. Biol. Chem. 226. 497-509.

Galili G, Amir R, Fernie AR. 2016. The Regulation of Essential Amino Acid Synthesis and Accumulation in Plants. Annu Rev Plant Biol. 67:153-78. doi: 10.1146/annurev-arplant-043015-112213.

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, et al. 2019. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat Genet. 51(6):1044–51.

Garcia I, Rodgers M, Pepin R, Hsieh T-F, Matringe M. 1999. Characterization and Subcellular Compartmentation of Recombinant 4-Hydroxyphenylpyruvate Dioxygenase from Arabidopsis in Transgenic Tobacco. Plant Physiol. 119(4):1507–16.

Gilliland LU, Magallanes-Lundback M, Hemming C, Supplee A, Koornneef M, Bentsink L, et al. 2006. Genetic basis for natural variation in seed vitamin E levels in Arabidopsis thaliana. Proc Natl Acad Sci USA .103(49):18834–41.

Hamilton JP, Sim S-C, Stoffel K, Van Deynze A, Buell CR, Francis DM. 2012. Single Nucleotide Polymorphism Discovery in Cultivated Tomato via Sequencing by Synthesis. Plant Genome J. 5(1):17.

Horvath A, Boikos S, Giatzakis C, Robinson-White A, Groussin L, Griffin KJ, et al. 2006. A genome-wide scan identifies mutations in the gene encoding phosphodiesterase 11A4 (PDE11A) in individuals with adrenocortical hyperplasia. Nat Genet. 38(7):794–800.

Huang BE, George AW. 2011. R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. Bioinformatics. 27(5):727–9.

Jin S, Daniell H. 2014. Expression of γ-tocopherol methyltransferase in chloroplasts results in massive proliferation of the inner envelope membrane and decreases susceptibility to salt and metal-induced oxidative stresses by reducing reactive oxygen species. Plant Biotechnol J. 12(9):1274–85.

Karunanandaa B, Qi Q, Hao M, Baszis SR, Jensen PK, Wong YH, et al. 2005. Metabolically engineered oilseed crops with enhanced seed tocopherol. 7:384–400.

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, et al. 2005. GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics. 21(8):1635–8.

Krieger-Liszkay A, Trebst A. 2006. Tocopherol is the scavenger of singlet oxygen produced by the triplet states of chlorophyll in the PSII reaction centre. In: Journal of Experimental Botany. 1677–84.

Lee BK, Kim SL, Kim KH, Yu SH, Lee SC, Zhang Z, et al. 2008. Seed specific expression of perilla ?-tocopherol methyltransferase gene increases a-tocopherol content in transgenic perilla (Perilla frutescens). Plant Cell Tissue Organ Cult. 92(1):47–54.

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. 2006. Gas chromatography mass spectrometry–based metabolite profiling in plants. Nat Protoc. 1(1):387–96.

Lu Y, Rijzaani H, Karcher D, Ruf S, Bock R. 2013. Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. Proc Natl Acad Sci. 110(8):623–32.

Mène-Saffrané L, Pellaud S. 2017. Current strategies for vitamin E biofortification of crops. Curr Opin Biotechnol. 44:189–97.

Miret JA, Munné-Bosch S. Redox signaling and stress tolerance in plants: A focus on vitamin E. Ann N Y Acad Sci. 2015;1340(1):29–38.

Munne-Bosch & Alegre. 2002. Interplay between ascorbic acid and lipophilic antioxidant defences in chloroplasts of water-stressed Arabidopsis plants. FEBS Lett. 524:145–8.

Muñoz P & Munné-Bosch S. 2019. Vitamin E in Plants: Biosynthesis, Transport, and Function. Trends Plant Sci. 24(11):1040-1051.

Nafati M, Cheniclet C, Hernould M, Do PT, Fernie AR, Chevalier C, Gévaudant F. The specific overexpression of a cyclin-dependent kinase inhibitor in tomato fruit mesocarp

cells uncouples endoreduplication and cell growth. Plant J. 2011 Feb;65(4):543-56. doi: 10.1111/j.1365-313X.2010.04446.x. Epub 2011 Jan 4. PMID: 21288265.

Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-3814. doi:10.1093/nar/gkg509

Nunes-Nesi A, Carrari F, Lytovchenko A, Smith AMO, Ehlers Loureiro M, Ratcliffe RG, et al. 2005. Enhanced Photosynthetic Performance and Growth as a Consequence of Decreasing Mitochondrial Malate Dehydrogenase Activity in Transgenic Tomato Plants. Plant Physiol. 137(2):611–22.

Osorio S, Alba R, Nikoloski Z, Kochevenko A, Fernie AR, Giovannoni JJ. 2012. Integrative Comparative Analyses of Transcript and Metabolite Profiles from Pepper and Tomato Ripening and Development Stages Uncovers Species-Specific Patterns of Network Regulatory Behavior. Plant Physiol. 159(4):1713–29.

Packer L, Weber SU, Rimbach G. 2001. Molecular Aspects of a-Tocotrienol Antioxidant Action and Cell Signalling. J Nutr. 131(2):369S-373S.

Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, Bouchet JP, et al. 2015. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnol J. 13(4):565-77. doi: 10.1111/pbi.12282.

Peh HY, Tan WSD, Liao W, Wong WSF. 2016. Vitamin E therapy beyond cancer: Tocopherol versus tocotrienol. Pharmacol Ther . 162:152–69.

Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, et al. 2014. Natural occurring epialleles determine vitamin e accumulation in tomato fruits. Nat Commun. 5, 4027. https://doi.org/10.1038/ncomms5027

Quadrana L, Almeida J, Otaiza SN, Duffy T, da Silva JVC, de Godoy F, et al. 2013. Transcriptional regulation of tocopherol biosynthesis in tomato. Plant Mol Biol. 81(3):309–25.

Ranc N, Muos S, Santoni S, Causse M. 2008. A clarified position for solanum lycopersicum var. cerasiforme in the evolutionary history of tomatoes (solanaceae). BMC Plant Biol. 20;8:130. doi: 10.1186/1471-2229-8-130.

Remington D, Thornsberry J. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA. 98(20): 11479–11484.

Rippert P, Scimemi C, Dubald M, Matringe M. 2004. Engineering Plant Shikimate Pathway for Production of Tocotrienol and Improving Herbicide Resistance. Plant Physiol. 134(1):92–100.

Roessner U, Willmitzer L, Fernie AR. 2001. High-Resolution Metabolic Phenotyping of Genetically and Environmentally Diverse Potato Tuber Systems. Identification of Phenocopies.Plant Physiol. 127(3):749–64.

Roessner-Tunali U, Hegemann B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, et al. 2003. Metabolic Profiling of Transgenic Tomato Plants Overexpressing Hexokinase Reveals That the Influence of Hexose Phosphorylation Diminishes during Fruit Development. Plant Physiol. 133(1):84–99.

Sadre R, Gruber J, Frentzen M. 2006. Characterization of homogentisate prenyltransferases involved in plastoquinone-9 and tocochromanol biosynthesis. FEBS Lett. 580(22):5357–62.

Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, et al. 2014. Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. Plant Physiol . 165(3):1120–32.

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, et al. 2008. Mode of Inheritance of Primary Metabolic Traits in Tomato. Plant Cell. 20(3):509–23.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet . 44(7):825–30.

Serbinova EA, Packer L. 1994. Antioxidant properties of α-tocopherol and α-tocotrienol. Methods Enzymol. 234:354-66.

Shintani D, DellaPenna D. 1998. Increasing Vitamin E content of plants through metabolic engineering. Science. 282(5396):2098–100.

Sim SC, van Deynze A, Stoffel K, Douches DS, Zarka D, Ganal MW, et al. 2012. High-Density SNP Genotyping of Tomato (Solanum lycopersicum L.) Reveals Patterns of Genetic Variation Due to Breeding. PLoS One. 7(9):1–18.

Sookwong P, Murata K, Nakagawa K, Shibata A, Kimura T, Yamaguchi M, et al. 2009. Cross-fertilization for enhancing tocotrienol biosynthesis in rice plants and QTL analysis of their F2 progenies. J Agric Food Chem. 57(11):4620–5.

Spicher L, Almeida J, Gutbrod K, Pipitone R, Dörmann P, Glauser G, et al. 2017. Essential role for phytol kinase and tocopherol in tolerance to combined light and temperature stress in tomato. J Exp Bot. 68(21–22):5845–56.

Stacey MG, Cahoon RE, Nguyen HT, Cui Y, Sato S, Nguyen CT, et al. 2016. Identification of Homogentisate Dioxygenase as a Target for Vitamin E Biofortification in Oilseeds. Plant Physiol. 172(3): 1506–1518.

Stahl E, Hartmann M, Scholten N, Zeier J. 2019. A Role for Tocopherol Biosynthesis in Arabidopsis Basal Immunity to Bacterial Infection.Plant Physiol. 181(3):1008–28.

Suarna C, Hood RL, Dean RT, Stocker R. 1993. Comparative antioxidant activity of tocotrienols and other natural lipid-soluble antioxidants in a homogeneous system, and in rat and human lipoproteins. Biochim Biophys Acta. 1166 (2-3):163-70.

Sussmann RAC, Fotoran WL, Kimura EA, Katzin AM. 2017. Plasmodium falciparum uses vitamin e to avoid oxidative stress. Parasites and Vectors. 10(1):1-8.

Suzuki YJ, Tsuchiya M, Wassall SR, Choo YM, Govil G, Kagan VE, et al. 1993. Structural and dynamic membrane properties of .alpha.-tocopherol and .alpha.-tocotrienol: Implication to the molecular mechanism of their antioxidant potency. Biochemistry. 32(40):10692–9.

Team. RDC. 2012. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Teng W, Li W, Zhang Q, Wu D, Zhao X, Li H, Han Y, Li W. 2017. Identification of quantitative trait loci underlying seed protein content of soybean including main, epistatic, and QTL × environment effects in different regions of Northeast China. 60(8):649-655. doi: 10.1139/gen-2016-0189.

Tian L, DellaPenna D, Dixon RA. 2007. The pds2 mutation is a lesion in the Arabidopsis homogentisate solanesyltransferase gene involved in plastoquinone biosynthesis. Planta. 226(4):1067–73.

Tzin V, Galili G. 2010. The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in Arabidopsis thaliana . Arabidopsis Book. 8: e0132.

Tzin V, Malitsky S, Aharoni A, Galili G. 2009. Expression of a bacterial bi-functional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in Arabidopsis. Plant J. 60(1):156–67.

Van Eenennaam AL, Lincoln K, Durrett TP, Valentin HE, Shewmaker CK, Thorne GM, et al. 2003. Engineering Vitamin E Content: From Arabidopsis Mutant to Soy Oil. Plant Cell. 15(12):3007–19.

Wang H, Xu S, Fan Y, Liu N, Zhan W, Liu H, et al. 2018. Beyond pathways: genetic dissection of tocopherol content in maize kernels by combining linkage and association analyses. Plant Biotechnol J. 16(8):1464-1475. doi: 10.1111/pbi.12889.

Wang X, Zhang C, Li L, Fritsche S, Endrigkeit J, Zhang W, et al. 2012. Unraveling the Genetic Basis of Seed Tocopherol Content and Composition in Rapeseed (Brassica napus L.). PLoS One. 7(11):1–18.

Warnes, Leisch. 2011. Package “genetics” Title Population Genetics.

Yusuf MA, Sarin NB. 2007. Antioxidant value addition in human diets: Genetic transformation of Brassica juncea with γ-TMT gene for increased α-tocopherol content. Transgenic Res. 16(1):109–13.

Zainal Z, Rahim AA, Radhakrishnan AK, Chang SK, Khaza'ai H. 2019. Investigation of the curative effects of palm vitamin E tocotrienols on autoimmune arthritis disease in vivo. Sci Rep. 9(1):16793. doi:10.1038/s41598-019-53424-7

Zhang C, Cahoon RE, Hunter SC, Chen M, Han J, Cahoon EB. 2013. Genetic and biochemical basis for alternative routes of tocotrienol biosynthesis for enhanced vitamin E antioxidant production. 628–39.

Zhang L, Luo Y, Zhu Y, Zhang L, Zhang W, Chen R, et al. 2013. GmTMT2a from soybean elevates the a-tocopherol content in corn and Arabidopsis. Transgenic Res. 22(5):1021–8.

Zhang W, Liu T, Ren G, Hörtensteiner S, Zhou Y, Cahoon EB, et al. Chlorophyll degradation: the tocopherol biosynthesis related phytol hydrolase in Arabidopsis seeds is still missing. Plant Physiol. 2014;166(1):70–9.

Zouine M, Maza E, Djari A, Lauvernier M, Frasse P, Smouni A, et al. TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. Plant J. 2017 ;92(4):727–35.

Figure and table legends

Figure 1. Phenotypic variability of tocochromanols accumulation in GWAS population of tomato fruits under two irrigation conditions. (a) Fruit tocochromanol average level reaction norms for each accession; 136 accessions of the GWAS population grown under control (light blue) and water deficit (pink) condition. Gray lines link genotype responses **(b)** ANOVA dissection of the total phenotypic variation for the GWAS panel showed as proportion of each effect in the total sum of squares (%SS): phenotype (P), pink stands for genotype (G), yellow stands for environment (E), light blue stands for the interaction (GxE) and grey stands for residual effects. Bellow each bar the significance of the p values is given; *** $p<0.001$, ** $p<0.01$, * $p<0.05$ and ns > 0.05.

Figure 2. GWAS and interval QTL mapping of tocochromanol contents in mature tomato fruits. 18 associations and 7 QTL were mapped by phenotyping a panel of 136 GWAS lines (Albert *et al.,* 2016) and 124 lines of a MAGIC population (Pascual *et al.,* 2015). Genomic regions significantly associated (p-value $<10^{-4}$) to variations in the different tocochromanol forms measured in fruits are depicted in light blue and pink for associations found with phenotypic data measured in plants cultivated under irrigation and drought respectively. Underline depicts QTL identified by interval mapping with the MAGIC population. Physical intervals (CI) were calculated based on the LD between significant and associated markers (see Experimental procedures). On the right of each chromosome candidate gene IDs (Table S3) are shown and their Mapman categories are indicated by different colors. α, δ, tot-t and α-t3 stand for α-tocopherol, δ-tocopherol total-tocopherol and α-tocotrienol, respectively. Gray box show Mapman categories by different reference colors and on the right significant enrichment of each category as ratio between QTL genes and total genes in the genome ($p<0.0001$, Fisher test).

Figure 3. Tocochromanol QTL mapping details and allelic effects. (a) Manhattan plots showing three significant (p-values $<10^{-4}$) associations with tot-t, α-t and α-t3 tocochromanol isoforms over genomic positions (x-axis) at the top of chromosomes 9 and bottoms of chromosomes 11 and 6, respectively. SNP markers with significant effects are depicted in red. Shadowed regions indicate the CI (calculated as described in Experimental procedures) for each QTL (see details in Table S3). Above each plot, all annotated genes (annotation version ITAG 2.4) within the corresponding CI are shown. Genes depicted in red types correspond to 4 of the identified candidate genes (see the complete list in Table S6). **(b)** Box-plots showing the allelic effects for the three associated markers: S09_04511795 (tot-t, "under drought conditions"), S11_53937168 (α-t, "under irrigation conditions") and S06_42560313 (α-t3, "under irrigation conditions"). Allelic effects for all detected associations are listed in Supplemental Table S2. **(c)** Genomic regions comprising α-t 11.1 and δ-t 12.1 QTL detected by IM with the MAGIC population. Shadowed regions indicate the CI. Above each plot 5 candidate genes (in red types) within the corresponding CI are shown. **(d)** Founder allelic effects estimated for α-t 11.1 and δ-t 12.1 QTL. For each QTL, the allelic effects were estimated as described in

Pascual *et al.* (2015) based on the founder haplotypes at QTL's position for each parental lines of the MAGIC population.

Figure 4. Engineering chorismate-tyrosine pathways in tomato fruits. **(a)** Schematic view of the pathways convergent to tocochromanol biosynthesis highlighting the structural genes identified as candidates in our genetic mapping**.** Gene IDs and abbreviations are detailed in Table S3. The PDH/HPPD by-passes are indicated in blue boxes. Abbreviations for intermediate metabolites are: glyceraldehyde 3-phosphate (GA3P); phosphoenol-pyruvate (PEP); Triptophane (Trp); Phenylalanine (Phe); Tyrosine (Tyr); geranylgeranyl-diphosphate (GeranylgeranylDP); chlorophyllide a (Chlla); phytyl-diphosphate (phytylDP); 2-methyl-6-phytyl-1,4-benzoquinol (MPBQ); 2,3-dimethyl-6-phytyl-1,4-benzoquinol (DMPBQ); 2-methyl-6-geranylgeranyl-1,4-benzoquinol (MGGBQ); 2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinol (DMGGBQ) and plastoquinol-9 (PQH2-9). **(b)** Representative HPLC-UV chromatograms of tocochromanol extracts from fruit samples (Breaker stage +4 days) of wild type (WT) and PH13 transgenic line (T0).

Figure 5. Expression levels of the heterologous transgenes and tocochromanol contents in chorismate–tyrosine pathway engineered T1 plants. (a) Relative *ScPDH* and *AtHPPD* genes transcript levels in fruits of 20 DPA, Breaker (Br) and Breaker+8 (red ripe) stages from three independent double-homozygous transgenic lines; PH13 (three siblings; #1, #2 and #5), PH17, PH30 and the wild-type control. Data are means ± SD of three biological replicates with two technical repeats normalized with the housekeeping *EIF4a* gene. **(b)** Tocochromanol contents in pericarp tissue from fruits at 20 DPA, Breaker (Br) and Breaker+8 (red ripe) stages for the same lines as in **(a)**. Data are means ± SD of three fruits from each transgenic line and controls. Asterisks indicate significant t-test differences (p<0.05) between WT and PH13 lines.

Figure 6. Chloroplast ultrastructure of tocochromanol accumulating plants. (a) Electron microscopic analysis of chloroplasts from tomato fruits at 20 DPA from three T1 progeny of the PH13 lines and wild-type plants showing the presence of plastoglobules (PG) as black spots in the proximity with thylakoids .**(b)** Number of plastoglobules

expressed per plastid area and **(c)** size of plastoglobules expressed as mean of a total of 60 chloroplasts for each transgenic line and control. Error bars indicate SD. Asterisks indicate significant differences between PH 13 and WT ($p<0.05$).

Figure 7. Expression analyses of structural genes of the tocochromanol-branching pathways and aromatic aminoacids profile in fruits of the PH transgenic lines. (a) Principal component analysis of relative expression data from 23 structural genes of the tocochromanol-related pathways assayed by the qPCR array platform in fruits of different developmental stages (20 DPA, breaker (BR) and (BR+8)) from three trangenics lines (PH13, PH17, PH30) and WT controls ($n \geq 3$). **(b)** Schematic view of the methyl erythritol (MEP), shikimate (SK), vitamin E-core and related pathways showing the relative expression levels for each individual gene measured as indicated at the color scale (-log2 ratio expression between PH13 and WT) in samples of 20 DPA, BR and BR+8 stages ($n \geq 3$). Tyrosine, phenylalanine and tryptophane were relatively quantified by GC-MS in the same samples. Significances were assayed by t-test (***$p<0.001$, ** $p<0.01$, * $p<0.05$). Enzyme encoding genes are named according to the following abbreviations: shikimate dehydrogenase/3-dehydroquinate dehydratase (DHQ); 4-Amino-4-deoxychorismate lyase (ADCL); Folate/biopterin transporter (fbt); anthranilate phosphoribosyltransferase (AnPrT); chorismate mutase (CM); prephenate aminotransferase (PAT); arogenate dehydrogenase (TyrA); tyrosine aminotransferase (TAT); L-tyrosine transporter (LAT1); 4-hydroxyphenylpyruvate dioxygenase (HPPD); Homogentisate 1,2-dioxygenase (HGO);geranylgeranyl pyrophosphate synthase (GGPS); geranylgeranyl reductase (GGDR); homogentisate phytyl transferase (VTE2); 2,3- dimethyl-5-phytylquinol methyltransferase (VTE3); tocopherol cyclase (VTE1); γ-tocopherol C-methyl transferase (VTE4). Paralogous genes are indicated by different numbers in brackets. Candidate genes found in the association mapping analyses are in dark pink font. Positive and negative feedback between aromatic aminoacids and enzymes are indicated with dash green arrows. The PDH/HPPD by-passes are indicated in blue boxes. Abbreviations for intermediate metabolites are as follow: glyceraldehyde 3-phosphate (GA3P); phosphoenol-pyruvate (PEP); Triptophane (Trp); Phenylalanine (Phe); Tyrosine (Tyr); geranylgeranyl-diphosphate (GeranylgeranylDP); chlorophyllide a (Chlla); phytyl-diphosphate (phytylDP); 2- methyl-6-phytyl-1,4-benzoquinol (MPBQ); 2,3-dimethyl-6-

phytyl-1,4-benzoquinol (DMPBQ); 2-methyl-6-geranylgeranyl-1,4-benzoquinol (MGGBQ); 2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinol (DMGGBQ).

Figure S1. Histogram plots of phenotypes measured in the populations used in this study. (a) Box-Cox transformed data of all 124 lines of the MAGIC population. Lambda values from data transformations are shown on the x-axis. **(b)** Box-Cox transformed data from all detected forms of tocochromanols measured in ripe fruits of 136 accessions of the GWAS population from plants grown under drought (pink) and irrigation(light blue) conditions. Lambda values from data transformations are shown on the x-axis.

Figure S2. Manhattan and QQplots of the associations detected after applying the MLMM of the experiments performed with fruits from plants cultivated under irrigation (I) and drought (D) conditions. Significantly associations (p-value<10^{-4}) for tocochromanol isoforms indicated in blue as: α, δ, and tot-t stand for α-tocopherol, δ-tocopherol, and total-tocopherol, respectively. α-t3: α-tocotrienol. Shadowed regions in yellow depict same markers associated with different tocochromanol isoforms.

Figure S3. Haplotype analyses of candidate genes detected by GWAS. Haplotype block and frequencies of non-synonymous substitutions are shown. Boxplots indicate the phenotypic values corresponding to the different phenotypic groups and significant effects ($p < 0.05$ Kruskal Wallis test) are shown by different letters. SNP positions on the corresponding gene scheme are indicated above each block. SNPs are coded as: nucleotide position from the first ATG_aminoacid substitution. Aminoacid substitutions indicated with red asterisks predicted to affect function estimated by SIFT (Ng & Henikoff, 2003). Blue arrows indicate SNP-defining clusters.

Figure S4. Constructs and expression analyses of *ScPDH* and *AtHPPD* transgenic plants at T0 generation. (a) Constructs used to transform tomato plants (*Solanum lycopersicum* cv. WVA106) by *Agrobacterium tumefaciens* as described in Experimental procedures section. **(b)** RT-PCR Analysis of the expression of the *ScPDH* and *AtHPPD* genes in fruits at 20 DPA stage of five regenerating lines (PH13, PH17, PH20, PH30 and PH26) and wildtype WT plants. Tomato *SlActin1* gene was used as internal control.

(c) Total tocopherols and tocotrienols contents indicated in gray and black bars respectively for PH13, PH17, PH30 lineas and WT (wild-type as control) in one replicate of T0 tomato fruits at stage Br+4 (orange) expressed in µg per g of dried fruit weight.

Figure S5. Accumulation pattern of tocochromanol during tomato fruit development and proportions expressed as α+β/γ+δ in fruits from PH13 lines and wild-type controls. (a) Total tocochromanols contents expressed as ug per g of dried fruits pericarp at division (5 and 10 DPA), expansion (20 and 30 DPA), ripening phases (Breaker, Br+4 (orange) and Br+8 (red ripe) for three
double homozygous sibling PH13 lines (T1) and the wild-type controls. **(b)** Contents of the different forms of tocochromanols (α-t, β-t, γ-t, δ-t and α-t3, γ-t3 and δ-t3) expressed as average of the three double homozygous sibling PH13 lines (T1) and the wild-type controls. Values are means ± SD. Asterisks indicate significant t-test differences between WT and PH13 lines; *** p<0.001, ** p<0.01, * p<0.05. **(c)** Tocopherol and **(d)** tocotrienol ratios (α+β/γ+δ) at division (5 and 10 DPA), expansion (20 and 30 DPA) and ripening phase (Breaker, Br+4 (orange) and Br+8 (red ripe)) for three three double homozygous sibling PH13 lines (T1) (in red color lines) and the wild-type controls (in red green lines). Pie charts of tocopherols **(e)** and tocotrienols **(f)** proportions at each developmental stage.

Figure S6. Levels of tocochromanols (tot-t and tot-t3) in source leaves from PH13 progeny lines and wild-type controls.

Figure S7. Linkage disequilibrium (R2) vs distances for each significant association marker of the GWAS associations detected after applying the MLMM. Physical distance of each confidence intervals (CI) were calculated based on the LD between significant (p-value $<10^{-4}$) and associated markers (see Experimental procedures) indicated in red numbers below X axis as base pairs. The 95th percentile of the unlinked-R2 distribution of markers equal to 0.28 was considered as the critical LD threshold (dashed lines in red). Then, for each significant marker, we computed LD with all the markers upstream and downstream on the same chromosome. Through plotting of R2 versus genetic distance between markers we defined both boundaries of the

confident interval as the intersection of the 'critical LD' threshold and the fitted curve of R2 regression. A nonlinear model described by Remington and Thornsberry (2001) was use to fitted the LD decay. The number of genes within each physical interval was identified from the tomato genome version SL2.50.

Tables

Table 1. Heritability, PVE and number of *loci* significantly associated with each trait detected with the GWAS panel.

Phenotype	Condition	h^2missing	h^2	PVE	No. of Associations
α-t	I	0.353	0.319	11.7	3
	D	-	0.307	-	0
α-t3	I	0.568	0.549	19.2	3
	D	0.600	0.612	15.9	3
β-t	I	-	0.548	-	0
	D	-	0.443	-	0
γ-t	I	-	0.472	-	0
	D	-	0.449	-	0
δ-t	I	0.343	0.368	14.2	2
	D	0.247	0.328	12.2	1
tot-t	I	0.210	0.306	10.4	2
	D	0.268	0.234	19.4	4

Global percentage of variation explained (PVE) by all the markers showing significant effects for a given trait are indicated under both watering regime (I: irrigated and D: drought conditions) from the optimal MLMM for each phenotype. Heritability values were estimated at step 0 by the MLMM (see Material and Methods). Missing heritability values (the percentage of the variance not explained by the markers) were calculated at the optimal step of the MLMM.

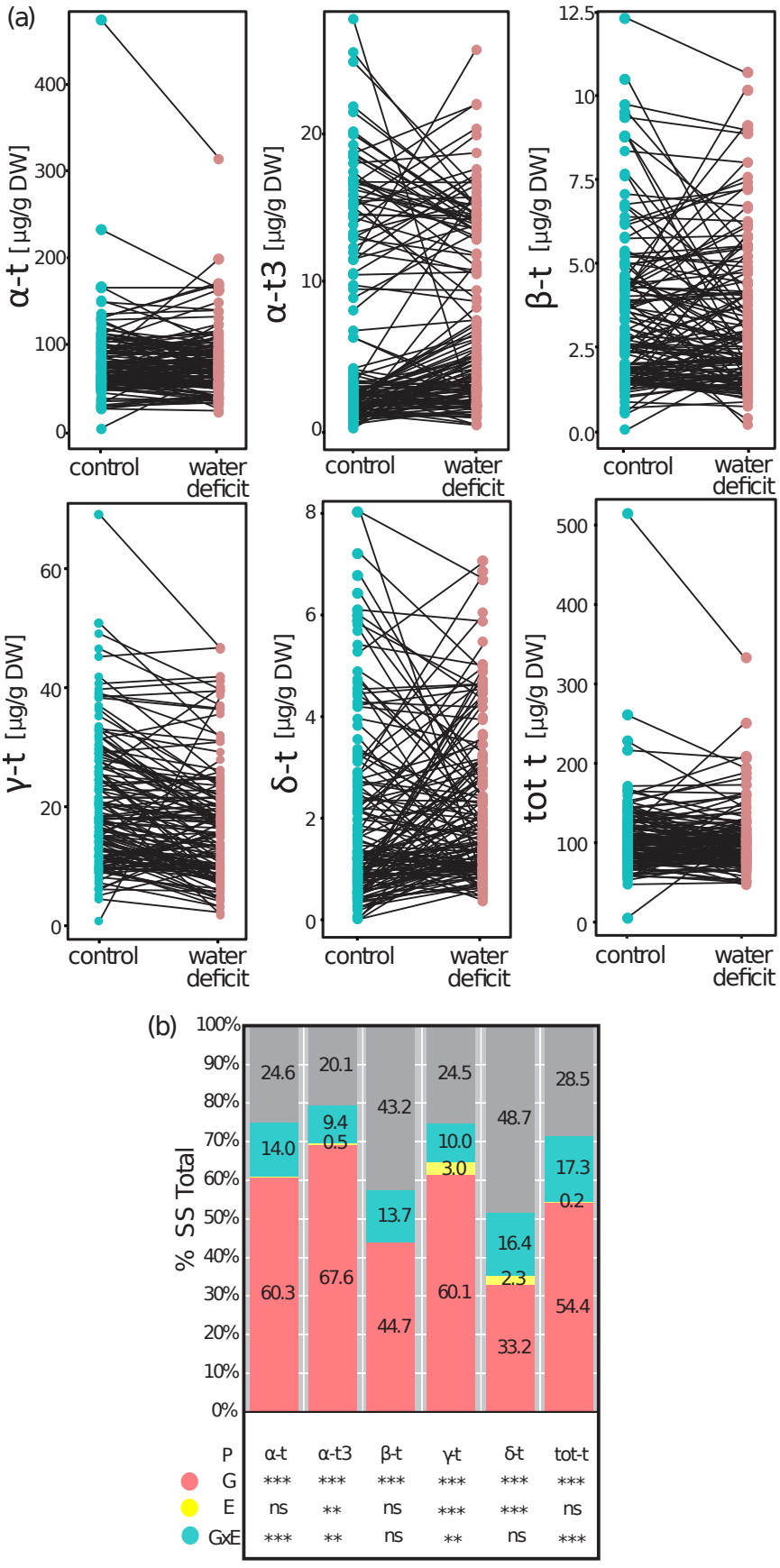

Figure 1. Phenotypic variability of tocochromanols accumulation in GWAS population of tomato fruits under two irrigation conditions. (a) Fruit tocochromanol average level reaction norms for each accession; 136 accessions of the GWAS population grown under control (light blue) and water deficit (pink) condition. Gray lines link genotype responses **(b)** ANOVA dissection of the total phenotypic variation for the GWAS panel showed as proportion of each effect in the total sum of squares (%SS): phenotype (P), pink stands for genotype (G), yellow stands for environment (E), light blue stands for the interaction (GxE) and grey stands for residual effects. Bellow each bar the significance of the p values is given; *** $p<0.001$, ** $p<0.01$, * $p<0.05$ and ns >0.05.

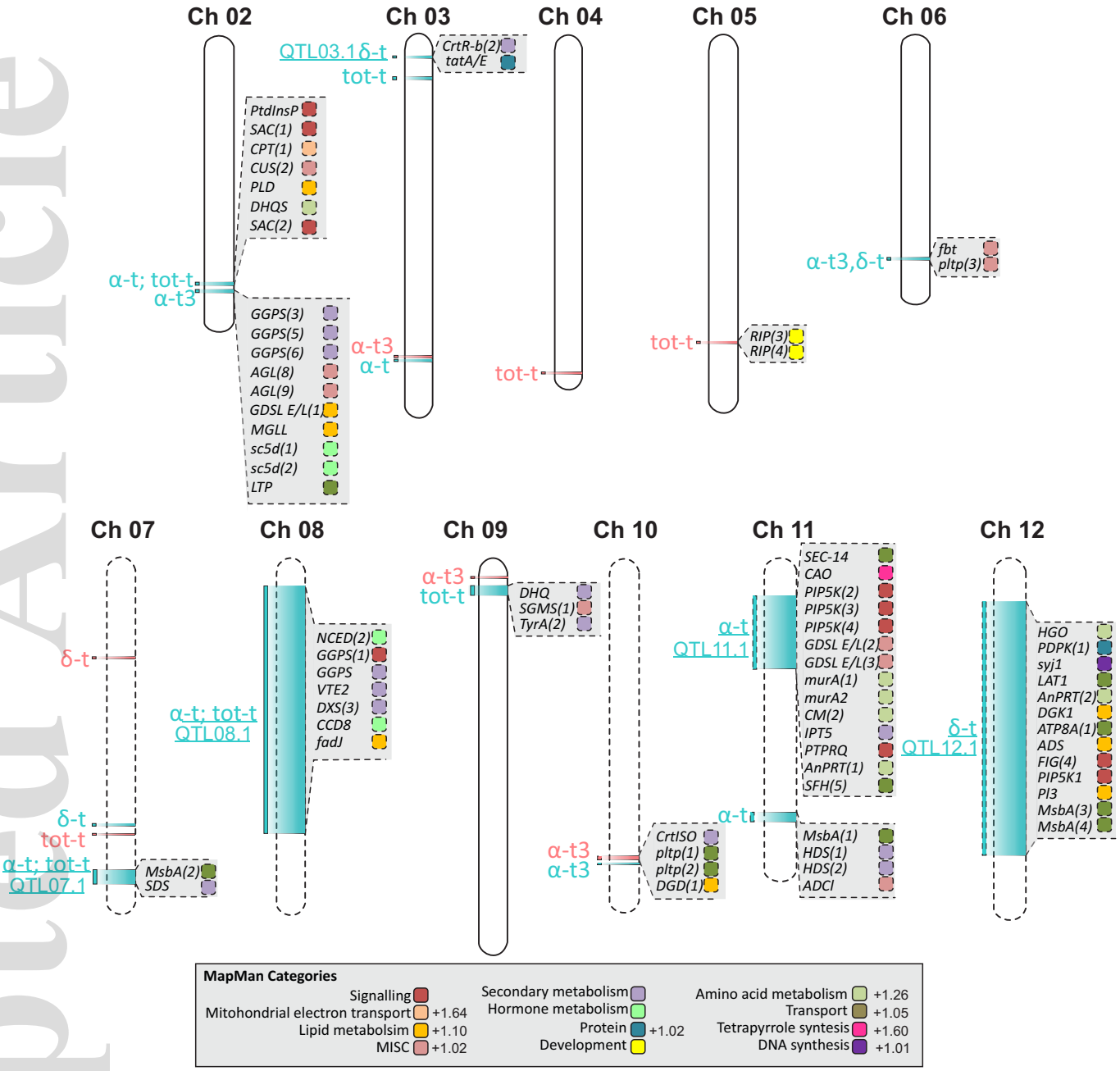

Figure 2. GWAS and interval QTL mapping of tocochromanol contents in mature tomato fruits. 18 associations and 7 QTL were mapped by phenotyping a panel of 136 GWAS lines (Albert et al., 2016) and 124 lines of a MAGIC population (Pascual et al., 2015). Genomic regions significantly associated (p-value $<10^{-4}$) to variations in the different tocochromanol forms measured in fruits are depicted in light blue and pink for associations found with phenotypic data measured in plants cultivated under irrigation and drought respectively. Underline depicts QTL identified by interval mapping with the MAGIC population. Physical intervals (CI) were calculated based on the LD between significant and associated markers (see Experimental procedures). On the right of each chromosome candidate gene IDs (Table S3) are shown and their Mapman categories are indicated by different colors. α, δ, tot-t and α-t3 stand for α-tocopherol, δ-tocopherol total-tocopherol and α-tocotrienol, respectively. Gray box show Mapman categories by different reference colors and on the right significant enrichment of each category as ratio between QTL genes and total genes in the genome (p<0.0001, Fisher test).

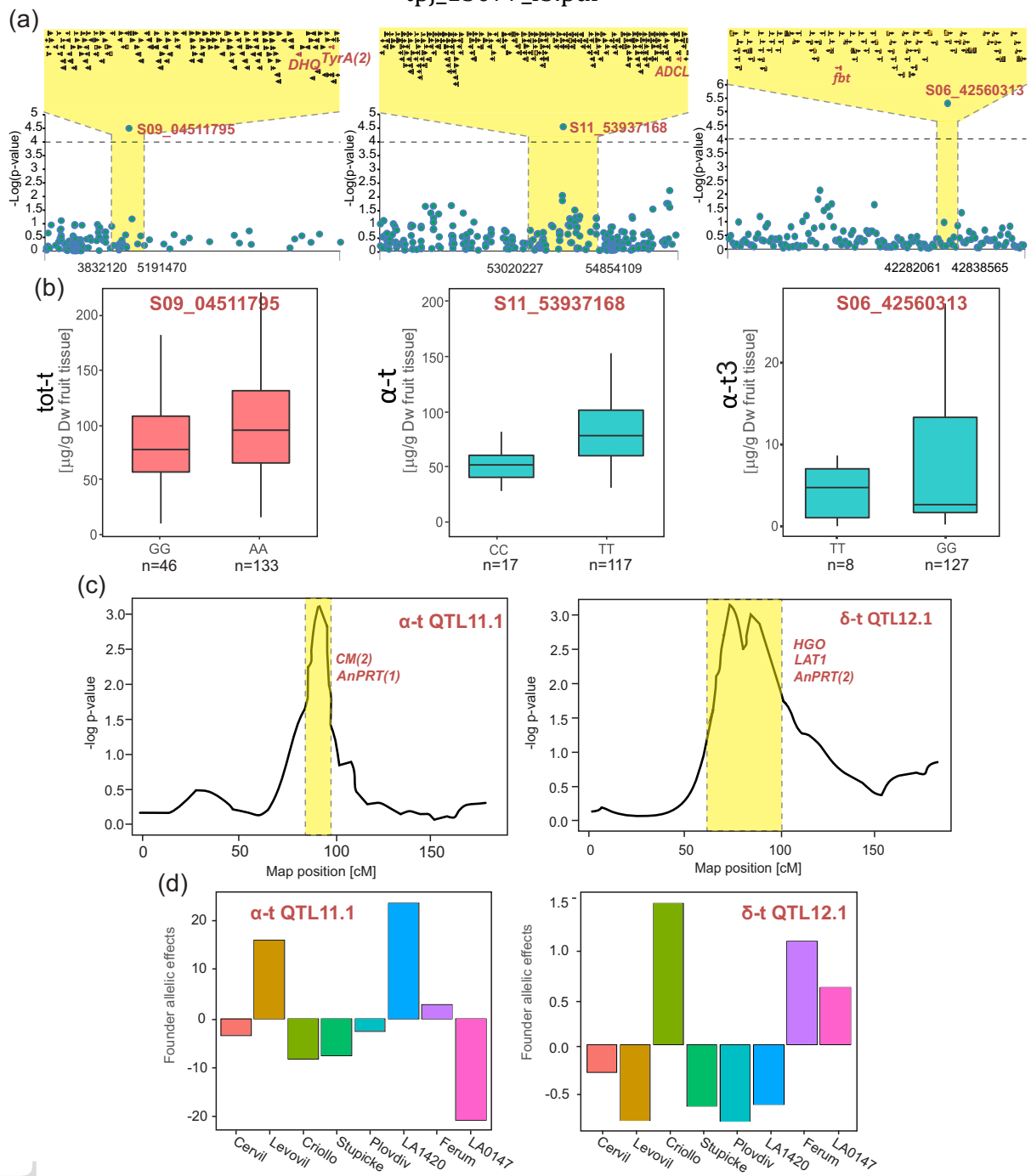

Figure 3. Tocochromanol QTL mapping details and allelic effects. (a) Manhattan plots showing three significant (p-values <10^{-4}) associations with tot-t, α-t and α-t3 tocochromanol isoforms over genomic positions (x-axis) at the top of chromosomes 9 and bottoms of chromosomes 11 and 6, respectively. SNP markers with significant effects are depicted in red. Shadowed regions indicate the CI (calculated as described in Experimental procedures) for each QTL (see details in Table S3). Above each plot, all annotated genes (annotation version ITAG 2.4) within the corresponding CI are shown. Genes depicted in red types correspond to 4 of the identified candidate genes (see the complete list in Table S6). **(b)** Box-plots showing the allelic effects for the three associated markers: S09_04511795 (tot-t, "under drought conditions"), S11_53937168 (α-t, "under irrigation conditions") and S06_42560313 (α-t3, "under irrigation conditions"). Allelic effects for all detected associations are listed in Supplemental Table S2. **(c)** Genomic regions comprising α-t 11.1 and δ-t 12.1 QTL detected by IM with the MAGIC population. Shadowed regions indicate the CI. Above each plot 5 candidate genes (in red types) within the corresponding CI are shown. **(d)** Founder allelic effects estimated for α-t 11.1 and δ-t 12.1 QTL. For each QTL, the allelic effects were estimated as described in Pascual et al. (2015) based on the founder haplotypes at QTL's position for each parental lines of the MAGIC population.

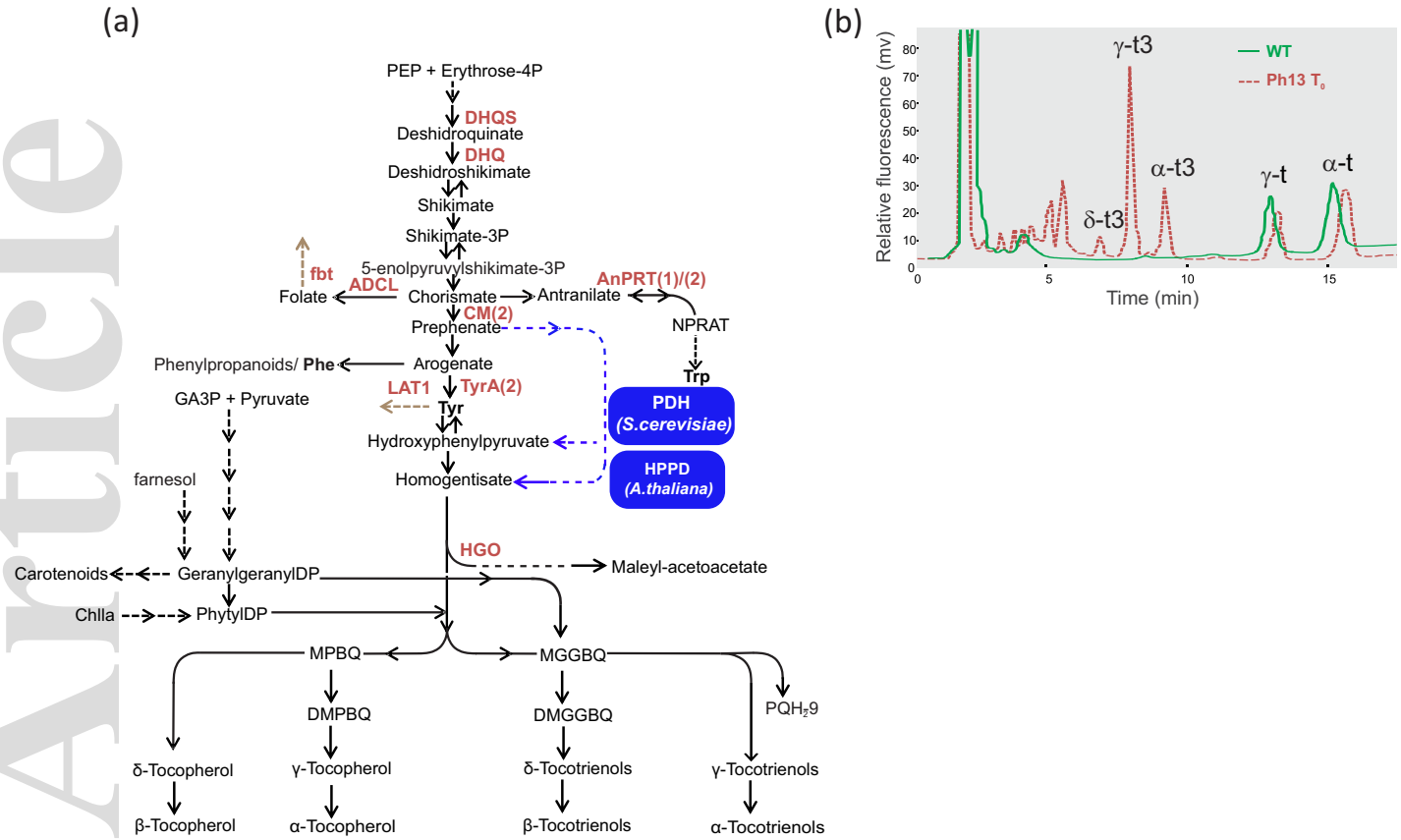

Figure 4. Engineering chorismate-tyrosine pathways in tomato fruits. **(a)** Schematic view of the pathways convergent to tocochromanol biosynthesis highlighting the structural genes identified as candidates in our genetic mapping. Gene IDs and abbreviations are detailed in Table S3. The PDH/HPPD by-passes are indicated in blue boxes. Abbreviations for intermediate metabolites are: glyceraldehyde 3-phosphate (GA3P); phosphoenol-pyruvate (PEP); Triptophane (Trp); Phenylalanine (Phe); Tyrosine (Tyr); geranylgeranyl-diphosphate (GeranylgeranylDP); chlorophyllide a (Chlla); phytyl-diphosphate (phytylDP); 2-methyl-6-phytyl-1,4-benzoquinol (MPBQ); 2,3-dimethyl-6-phytyl-1,4-benzoquinol (DMPBQ); 2-methyl-6-geranylgeranyl-1,4-benzoquinol (MGGBQ); 2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinol (DMGGBQ) and plastoquinol-9 (PQH2-9). **(b)** Representative HPLC-UV chromatograms of tocochromanol extracts from fruit samples (Breaker stage +4 days) of wild type (WT) and PH13 transgenic line (T0).

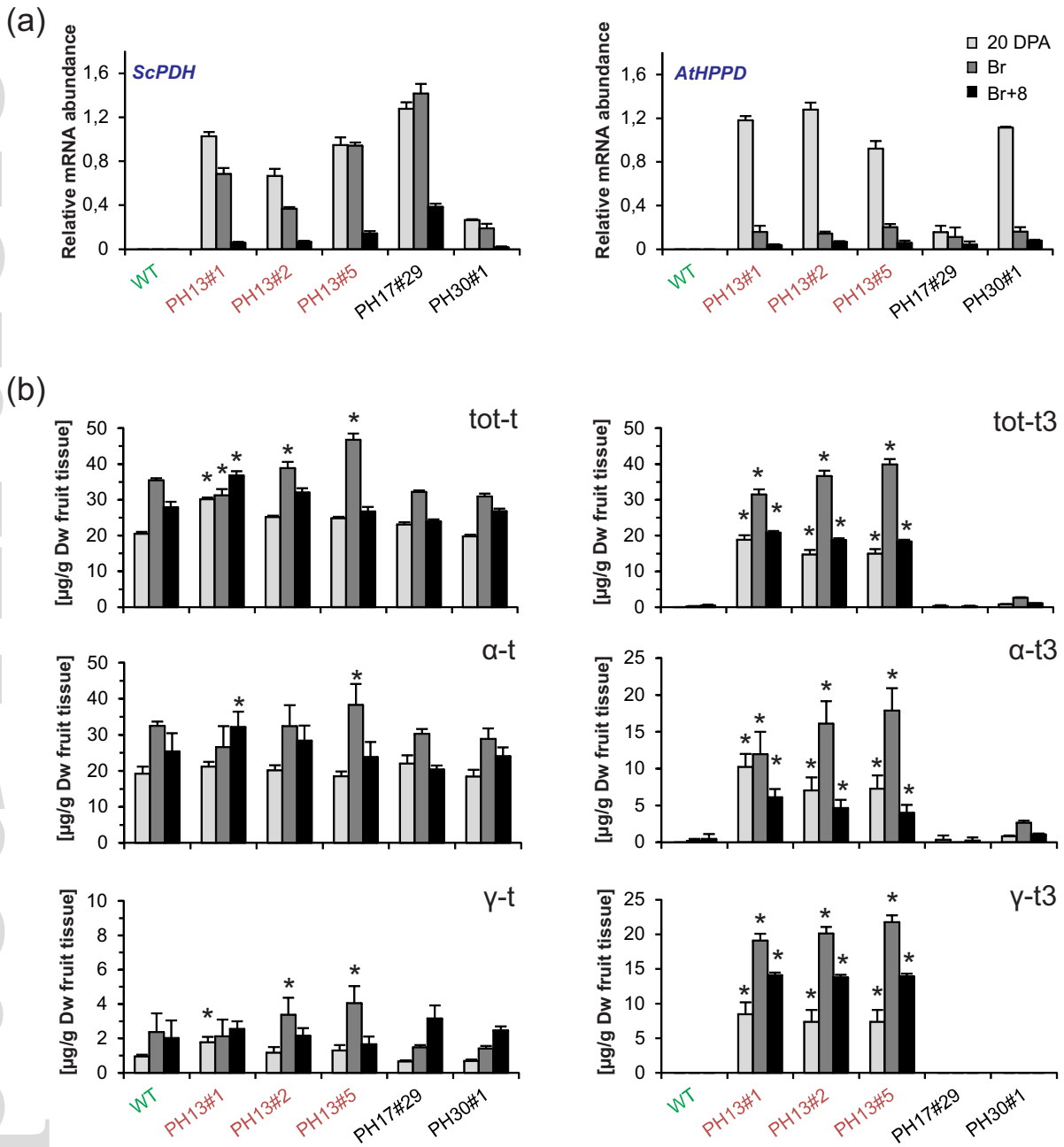

Figure 5. Expression levels of the heterologous transgenes and tocochromanol contents in chorismate–tyrosine pathway engineered T1 plants. **(a)** Relative *ScPDH* and *AtHPPD* genes transcript levels in fruits of 20 DPA, Breaker (Br) and Breaker+8 (red ripe) stages from three independent double-homozygous transgenic lines; PH13 (three siblings; #1, #2 and #5), PH17, PH30 and the wild-type control. Data are means ± SD of three biological replicates with two technical repeats normalized with the housekeeping *EIF4a* gene. (b) Tocochromanol contents in pericarp tissue from fruits at 20 DPA, Breaker (Br) and Breaker+8 (red ripe) stages for the same lines as in **(a)**. Data are means ± SD of three fruits from each transgenic line and controls. Asterisks indicate significant t-test differences ($p<0.05$) between WT and PH13 lines.

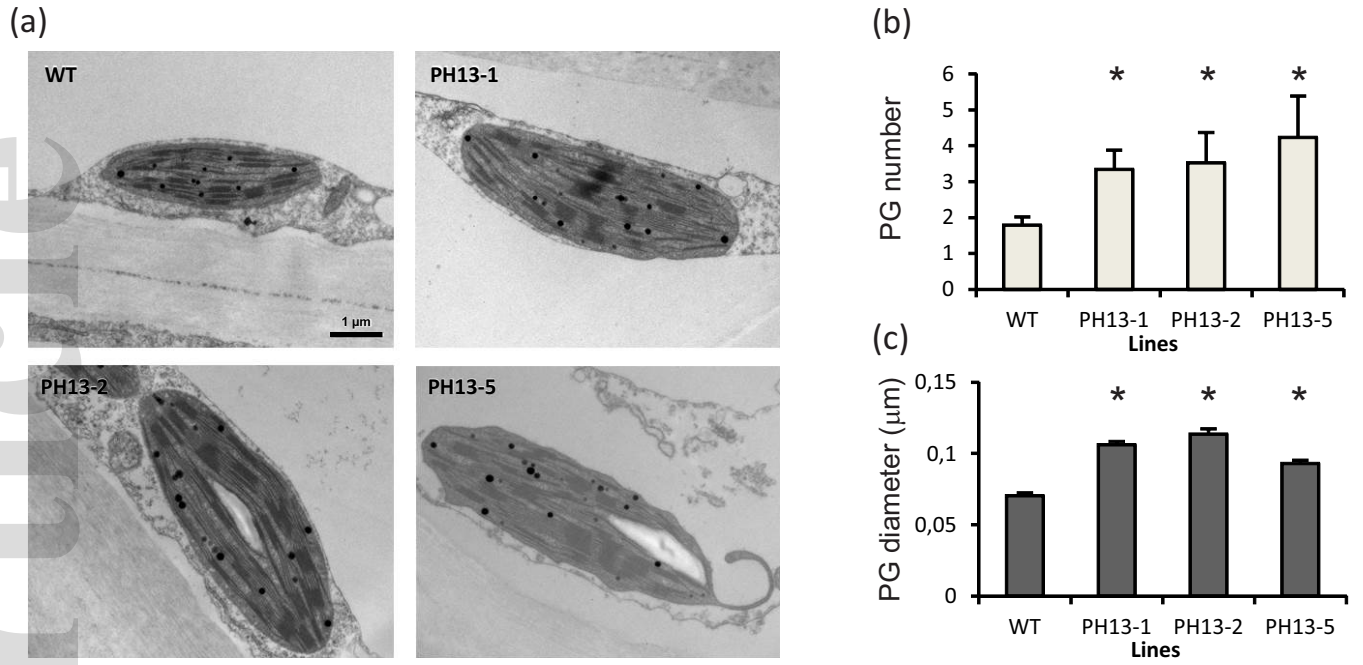

Figure 6. Chloroplast ultrastructure of tocochromanol accumulating plants. (a) Electron microscopic analysis of chloroplasts from tomato fruits at 20 DPA from three T1 progeny of the PH13 lines and wild-type plants showing the presence of plastoglobules (PG) as black spots in the proximity with thylakoids .**(b)** Number of plastoglobules expressed per plastid area and **(c)** size of plastoglobules expressed as mean of a total of 60 chloroplasts for each transgenic line and control. Error bars indicate SD. Asterisks indicate significant differences between PH 13 and WT ($p<0.05$).

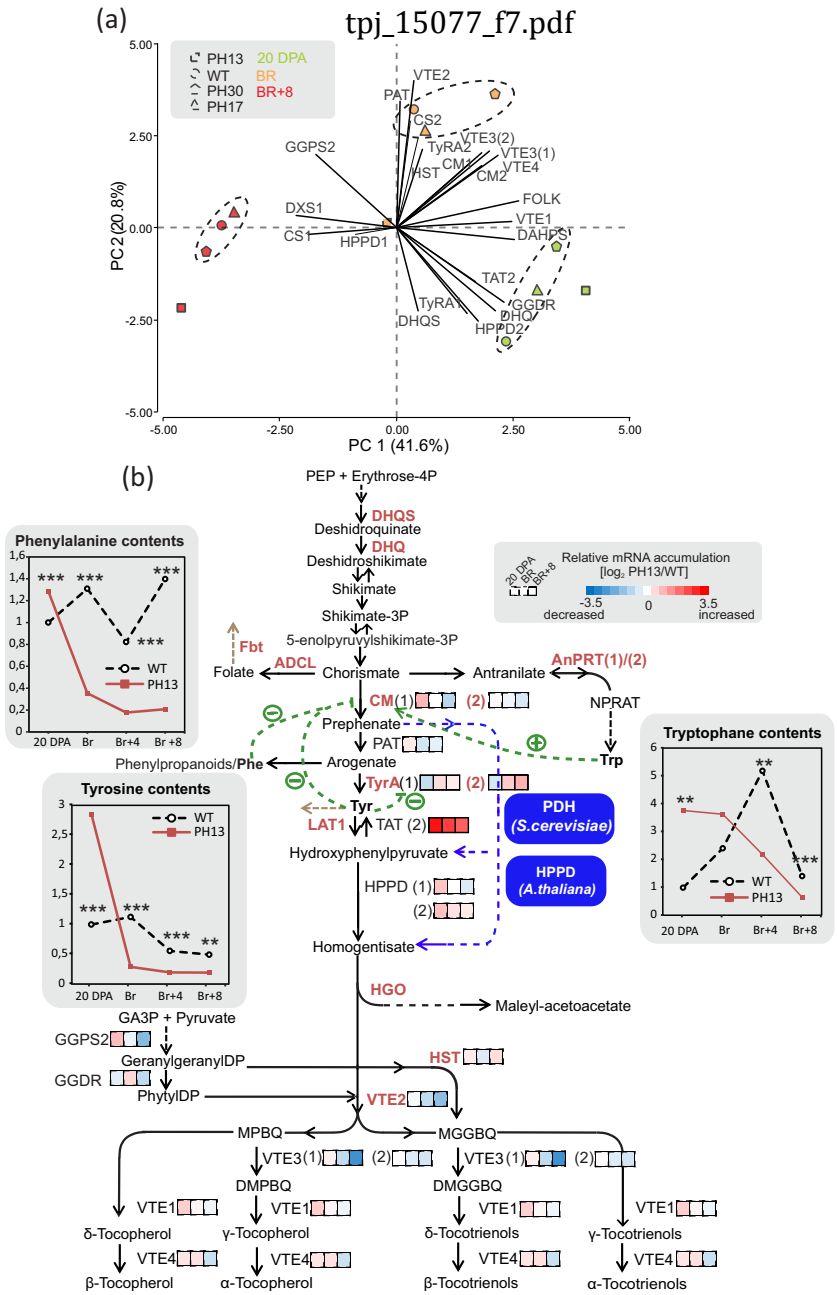

Figure 7. Expression analyses of structural genes of the tocochromanol-branching pathways and aromatic aminoacids profile in fruits of the PH transgenic lines. (a) Principal component analysis of relative expression data from 23 structural genes of the tocochromanol-related pathways assayed by the qPCR array platform in fruits of different developmental stages (20 DPA, breaker (BR) and (BR+8)) from three trangenics lines (PH13, PH17, PH30) and WT controls (n ≥ 3). **(b)** Schematic view of the methyl erythritol (MEP), shikimate (SK), vitamin E-core and related pathways showing the relative expression levels for each individual gene measured as indicated at the color scale (-log2 ratio expression between PH13 and WT) in samples of 20 DPA, BR and BR+8 stages (n ≥ 3). Tyrosine, phenylalanine and tryptophane were relatively quantified by GC-MS in the same samples. Significances were assayed by t-test (***p<0.001, ** p<0.01, * p<0.05). Enzyme encoding genes are named according to the following abbreviations: shikimate dehydrogenase/3-dehydroquinate dehydratase (DHQ); 4-Amino-4-deoxychorismate lyase (ADCL); Folate/biopterin transporter (fbt); anthranilate phosphoribosyltransferase (AnPrT); chorismate mutase (CM); prephenate aminotransferase (PAT); arogenate dehydrogenase (TyrA); tyrosine aminotransferase (TAT); L-tyrosine transporter (LAT1); 4-hydroxyphenylpyruvate dioxygenase (HPPD); Homogentisate 1,2-dioxygenase (HGO);geranylgeranyl pyrophosphate synthase (GGPS); geranylgeranyl reductase (GGDR); homogentisate phytyl transferase (VTE2); 2,3- dimethyl-5-phytylquinol methyltransferase (VTE3); tocopherol cyclase (VTE1); γ-tocopherol C-methyl transferase (VTE4). Paralogous genes are indicated by different numbers in brackets. Candidate genes found in the association mapping analyses are in dark pink font. Positive and negative feedback between aromatic aminoacids and enzymes are indicated with dash green arrows. The PDH/HPPD by-passes are indicated in blue boxes. Abbreviations for intermediate metabolites are as follow: glyceraldehyde 3-phosphate (GA3P); phosphoenol-pyruvate (PEP); Triptophane (Trp); Phenylalanine (Phe); Tyrosine (Tyr); geranylgeranyl-diphosphate (GeranylgeranylDP); chlorophyllide a (Chlla); phytyl-diphosphate (phytylDP); 2- methyl-6-phytyl-1,4-benzoquinol (MPBQ); 2,3-dimethyl-6-phytyl-1,4-benzoquinol (DMPBQ); 2-methyl-6-geranylgeranyl-1,4-benzoquinol (MGGBQ); 2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinol (DMGGBQ).